Received 9 October 2024; accepted 10 November 2024. Date of publication 14 November 2024; date of current version 27 November 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3499370

The Chronicles of 5G Non-Standalone: An Empirical Analysis of Performance and Service Evolution

GIUSEPPE CASO [1] (Member, IEEE), MOHAMMAD RAJIULLAH [1], KONSTANTINOS KOUSIAS [2], USMAN ALI [3] (Student Member, IEEE), NADIR BOUZAR [3], LUCA DE NARDIS [3] (Member, IEEE), ANNA BRUNSTROM [1,4] (Member, IEEE), OZGU ALAY [1,2] (Member, IEEE), MARCO NERI [5], AND MARIA-GABRIELLA DI BENEDETTO [3] (Fellow, IEEE)

[1] Department of Mathematics and Computer Science, Karlstad University, 651 88 Karlstad, Sweden
[2] Department of Informatics, University of Oslo, 0373 Oslo, Norway
[3] Department of Information Engineering, Electronics and Telecommunication, Sapienza University of Rome, 00184 Rome, Italy
[4] Institute of Software Engineering and Technologies, University of Malaga, 29071 Malaga, Spain
[5] Rohde & Schwarz, Mobile Network Testing Unit, 00156 Rome, Italy

CORRESPONDING AUTHOR: G. CASO (e-mail: giuseppe.caso@kau.se)

This work was supported in part by the Knowledge Foundation of Sweden; in part by the European Union through the 6G SNS Programme under Grant 101139172 (6G-PATH) and Grant 101096452 (IMAGINE-B5G); and in part by the European Union—Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP B53C22004050001, partnership on "Telecommunications of the Future" (Program "RESTART") under Grant PE00000001.

ABSTRACT Fifth Generation (5G) systems have been commercially available worldwide for at least a couple of years, with mid-band Non-Standalone (NSA) being the deployment mode preferred by Mobile Network Operators (MNOs). Empirical analyses have provided so far key insights on 5G NSA performance from different perspectives, but most of these works consider short time periods to drive conclusions. In this paper, we investigate the evolution of 5G NSA considering deployment, performance, and services, including positioning. We perform a large-scale measurement campaign in two phases (2021 and 2023), covering six MNOs in two European countries, Italy and Sweden. Our results show significant differences in network deployment and performance, with increasing network density and frequencies but, at times, decreasing downlink throughput performance. For the latter, we identify worse radio coverage and connectivity issues as root causes. By using a standardized methodology, we also evaluate the performance of new services such as real-time gaming and augmented/virtual reality, and reveal that stable 5G connectivity is key to meet their requirements. Similarly, we highlight the negative effects of roaming on performance. Finally, we evaluate 5G fingerprinting positioning systems and show that a higher accuracy is achievable in denser 5G deployments.

INDEX TERMS 5G mobile systems, large-scale measurements and analyses, performance and service evolution, user positioning.

I. INTRODUCTION

TO ADDRESS increasingly stringent Quality of Service (QoS) and Quality of Experience (QoE) requirements, e.g., throughput and latency/reliability demands of enhanced Mobile Broadband (eMBB) and Ultra-Reliable Low Latency Communication (URLLC), Fifth Generation (5G) networks are designed and deployed aside Fourth Generation (4G), e.g., Long Term Evolution (LTE) and LTE-Advanced (LTE-A) systems by the 3rd Generation Partnership Project (3GPP).

3GPP has standardized 5G New Radio (NR) in Release 15 (Rel-15), along with two deployment modes, Non-Standalone (NSA) and Standalone (SA). Both use a NR Radio Access Network (RAN), formed by Next Generation Node Bs (gNBs) and Physical Cell IDs (PCIs), which may operate in low (< 1 GHz), mid (1-7 GHz), and/or high (> 24 GHz) frequency bands.

5G NSA allows a Mobile Network Operator (MNO) to use the 4G Core Network (CN), while 5G SA requires a 5G CN [1]. For initial deployments, 5G NSA is thus a less costly solution but introduces tight 4G-5G inter-working and specific challenges compared to 5G SA, e.g., in terms of configurations and procedures.

At the time of writing (October 2024), several European, Asian, and U.S. MNOs deployed 5G NSA networks in low/mid bands, with a few 5G SA deployments, using also high bands, active in the U.S. and more recently in India and Germany [2], [3]. This ongoing deployment allows for empirical research on system and performance, which is key for better understanding such complex systems, identifying potential bottlenecks, and designing enhanced solutions for management and QoS/QoE performance, also based on Artificial Intelligence (AI) and Machine Learning (ML) [4].

Considering that 5G NSA networks have now been operational for at least a couple of years in many areas, it is important to start analyzing them from an *evolution* perspective. In this paper, we use the term *evolution* in its more general connotation, i.e., *"a gradual process of change and development"*, not directly implying enhancements and/or improvements, e.g., of a system and/or its performance/conditions. Therefore, our main goal is to empirically assess how 5G NSA networks are evolving (i.e., changing) over a long time span (i.e., years) in terms of aspects including deployment, performance, and services. As a matter of fact, as further detailed in Section II, empirical research has so far mostly focused on shorter time analyses, providing key insights but with no discussions on long-term changes. Moreover, the analysis of *beyond-eMBB* services has been mostly carried out with simple methodologies (e.g., ping), which do not provide realistic assessments and neglect QoE evaluations. Finally, 5G-based user positioning, which is receiving increasing attention for enabling location-based services, has been unexplored on commercial networks, with analyses carried out on simulated data or in testbeds [5].

In this paper, we move a step beyond the state of the art and provide an in-depth, longitudinal investigation of 5G NSA networks from deployment, performance, and service perspectives. To do so, we leverage our first large-scale measurement campaign, executed in 2020-2021 and open-sourced in [4], and perform a new large-scale measurement campaign in 2023, with the corresponding dataset described in this paper and also open-sourced.[1] Across campaigns, we cover the 5G NSA networks of 6 MNOs in 3 cities (Rome, Stockholm, and Karlstad) of 2 European countries (Italy and Sweden), in indoor/outdoor and static/mobile scenarios. Our analysis shows several interesting insights, summarized along with our main contributions as follows:

1) *Network Deployment and Throughput Performance (Section IV):* We study how 5G NSA deployments have changed over time and observe clear growing trends, with an increasing number of deployed PCIs and a broader use of multiple frequency bands, with some of the MNOs under study operating in both low and mid bands in 2023. We further study the change over time of downlink (DL) and uplink (UL) throughput performance. On the one hand, we observe a significant decrease in DL throughput in 2023. Upon deeper examination, we identify two main factors contributing to this decline: i) inferior radio coverage and ii) 5G connectivity issues. On the other hand, we also observe performance gains in the UL throughput of one MNO, which highlights the benefits achievable through an efficient utilization of 4G-5G Dual Connectivity (DC) in UL.
2) *Interactive Services and Roaming Performance (Section V):* We expand our analysis to include interactive applications like real-time gaming and Augmented/Virtual Reality (AR/VR), thus examining the performance of beyond-eMBB services. Unlike existing literature, we use a standardized approach to assess the QoS/QoE. Our findings indicate that maintaining stable 5G connectivity is crucial for meeting the demands of these services. Current 5G deployments face challenges with data-intensive AR/VR applications, suggesting the necessity for alternative architectural solutions. Furthermore, our analysis encompasses international roaming, revealing that the detrimental effect on latency of home-routed roaming persists in 5G networks. The fallback to 4G observed for MNOs that do not support 5G roaming highlights additional roaming performance issues.
3) *Handover and Connectivity Management (Section VI):* We investigate how the configurations for handover (HO) and connectivity management impact performance. In 2023, we observe a more dynamic and aggressive use of the 5G RAN across the MNOs under study, demonstrating a higher willingness to let 5G-capable devices use the 5G RAN even in worse coverage conditions. Our results reveal that this may hinder 5G connectivity, resulting in significantly lower DL throughput and performance degradation for real-time gaming services.
4) *Positioning (Section VII):* Considering that precise user position is crucial for many emerging services, we experimentally validate 5G fingerprinting positioning systems. Our results show that the proposed approach provides accurate positioning, and the evolution towards denser and multi-frequency 5G networks brings remarkable benefits to positioning accuracy.
5) *Dataset:* Along with the data description in the paper, we open-source our dataset for further investigation by the research community.

II. RELATED WORK

In recent years, several studies have empirically analyzed different technology and performance aspects of 5G commercial networks. Depending on the deployment choices of the MNOs in the region under study, the analyses targeted 5G NSA and/or SA deployments in different bands.

[1]The portion of the 2023 dataset used in this paper is available at https://doi.org/10.5281/zenodo.14073310.

Initial work was carried out in [6], [7] for the U.S. and [8] for China, which investigated throughput/latency performance, the impact of coverage, deployment strategies, mobility management, and server location on performance, and the power consumption of the 5G User Equipment (UE). For Europe, and specifically for Italy, we provided similar characterizations in [9], [10], [11], by using the dataset collected in 2020-2021 and disclosed in [4], which we also partially use in the present paper for the time evolution analyses. Moreover, by exploiting a single measurement collected in 2023, our poster paper in [12] preliminary highlighted the importance of longitudinal studies, such as the one carried out in the present contribution.

Moving from physical (PHY) to higher layers, we characterized outdoor-to-indoor propagation of mid-band NSA deployments in [13], [14], while the authors of [15] applied ML for predicting mid-band and high-band coverage. The work in [16] analyzed PHY latency in high-band NSA deployments and discussed factors affecting end-to-end latency, including server location, HOs, and UE energy-saving mechanisms. At higher layers, a study of throughput in high-band NSA deployments was carried out in [17], resulting in a ML-based throughput prediction scheme. HOs in NSA and (partially) SA deployments were analyzed in [18], where a HO prediction scheme was also proposed. Similarly, [19] focused on HO (mis-)configurations at Radio Resource Control (RRC) layer, while the use of the Rel-16 conditional HO mechanism was discussed in [20] for public transportation scenarios. At the application layer, [21] used throughput traces collected in high-band 5G deployments to conduct trace-driven simulations, ultimately showing how data-hungry applications (e.g., volumetric video streaming) can exploit the high throughput offered by 5G and overcome undesired yet continuous fluctuations.

Further large-scale data-driven analyses were recently carried out in [22], [23], [24], [25], [26]. Apart from [22], which covered different cities and MNOs in Oman and showed how the ongoing transition towards mid-band NSA deployments could lead to throughput and latency enhancements, the other works were all performed in the U.S. In particular, [23] analyzed coverage, throughput, and latency performance in low, mid, and high bands, finding that throughput and latency performance were comparable between mid-band 4G and 5G NSA/SA networks, with high-band 5G leading to the best throughput although limited by body interference, obstructions, and UE overheating. NSA and SA deployments were compared in [24], which showed SA often outperforming NSA in terms of throughput and latency, but with limitations requiring further optimization. The 5G potential of providing high throughput was also investigated in [25], which detailed a set of performance issues, analyzed the root causes, and proposed a fix called 5GBoost. Among other issues, it was observed that 5G was still underutilized due to policies that negatively affected performance, including slow multi-round RRC configurations. Low 5G coverage and often poor user performance were also observed in [26], where a cross-continental trip from East to West Coast in the U.S. highlighted the impact on throughput and latency of several factors, including coverage, user speed, and server location.

All the above works were based on measurements performed during within-year periods (up to 10 months [25]), thus not covering longer time spans as we do in our work through data collected in 2020-2021 and in 2023. Moreover, they were mostly focused on investigating throughput and latency performance, via proprietary Speedtests and/or open-source tools like iPerf and ping (for latency). Some recent work has also started to shed light on how services provided by 5G networks are rapidly growing, e.g., from MNO [27] and content provider [28] perspectives. This emphasizes the importance of studying 5G systems also from a service perspective, towards unveiling both potentials and bottlenecks in achieving the QoS/QoE requirements of beyond-eMBB services. However, so far, service-dedicated testing was mostly focused on eMBB, e.g., Web browsing and video conferencing/streaming [6], [7], [8], [18], with eMBB-URLLC services, including real-time cloud gaming, AR, and autonomous driving preliminarily tested only recently [9], [10], [11], [12], [18], [26]. More effort is needed in this direction, also considering that in-house developed tests, e.g., as in [26], are of great value but do not help towards wide comparison and repeatability. In this context, our work moves a step beyond by using a systematic and standardized QoS/QoE evaluation for beyond-eMBB services, as detailed in the next sections. Finally, very preliminary analyses were dedicated to 5G-based positioning, as surveyed in [5]. Apart from an initial characterization we carried out in [4], the present work provides the first in-depth empirical investigation of 5G positioning systems based on the fingerprinting technique, disclosing several insights on the main factors determining positioning accuracy and how accuracy was affected by the time evolution of 5G network deployments.

III. SETUP, METHODOLOGY, AND DATASET

In this section, we present the setup and methodology used during our measurements, and describe the collected dataset. Aiming at performing valid comparative analyses, we adopted the same setup (i.e., hardware and software) and methodology (i.e., configurations and tests) for our collections in 2020-2021 and 2023, as further detailed next.

For simplicity, in the following we will refer to the data collection in Rome, Italy, between December 2020 and January 2021, as the 1st collection phase, and to the collection in Rome, Italy, and in Stockholm and Karlstad, Sweden, between March 2023 and November 2023, as the 2nd collection phase. Although information on the 1st collection phase can be found in [4], we also report it in this paper for a direct comparison with the 2nd collection phase.

A. SETUP

During the 1st collection phase, we used a setup including an omnidirectional Radio Frequency (RF) antenna operating in

FIGURE 1. Measurement setup used during the 2nd collection phase: i) RF antenna, ii) GPS antenna, iii) R&S TSMA6, and iv) two 5G-capable UEs. A tablet connected via Wireless Local Area Network (WLAN) was used for accessing the software running in the R&S TSMA6, including ROMES.

the 698-3800 MHz frequency range, a synchronized Global Positioning System (GPS) antenna for geo-mapping our measurements, a 5G-capable UE (Samsung S20) embedded with the Rohde & Schwarz (R&S) Qualipoc Android app [29], and the R&S TSMA6, a system formed by a spectrum scanner and an Intel Windows PC. The latter runs ROMES, an R&S software enabling measurement configuration and inspection, and post-collection data exporting.

During the 2nd collection phase, we used the same setup but with two UEs used simultaneously, as shown in Fig. 1. In both phases, we used SIM cards from different MNOs to execute the performance tests.

B. METHODOLOGY

Our methodology comprised the parallel execution of passive network monitoring and active performance testing, under different scenarios identified in several locations and areas in the three cities under test.

1) SCENARIOS

The collection phases were organized in sub-campaigns in three main scenarios: i) Indoor Static (IS), for static measurements executed at different indoor locations; ii) Outdoor Driving (OD), for measurements executed while driving a car in Italy or being on a bus in Sweden; and iii) Outdoor Walking (OW), for measurements executed while walking around the cities. To increase the statistical significance of our measurements, we repeated the data collection in each location/area multiple times over different days and times of the day, with a larger number of repetitions per location/area during the 2nd collection phase.

In this paper, we focus on IS and OD scenarios and thus only describe these two in the following, considering that, during both collection phases, we executed a higher number of measurements in common IS locations and OD areas, which allows for the longitudinal analysis reported in the next sections. Moreover, we refer to both IS locations and OD areas simply as *locations*.

As reported in [4], during the 1st collection phase, we collected data in 10 IS and 6 OD locations in Rome. In the same city, during the 2nd collection phase, we covered 9 IS and 4 OD locations. The total length of paths covered in OD locations in Rome was 36.5 km during the 1st collection phase and 26.2 km during the 2nd collection phase, with more than 800 km covered across phases when also considering the repetitions in the OD locations. 7 IS locations (offices at the 2nd floor of the Department of Information Engineering, Electronics and Telecommunications (DIET) of Sapienza University of Rome) and 1 OD location (a long circular path of about 3 km) were in common across phases, thus forming the basis for our comparisons and analyses over time. Such locations were either in the historical city center (IS) or in a dense urban area (OD), and experienced no landscape changes across phases (e.g., no significant furniture changes and/or new buildings in the surrounding areas). This consistency reinforces the representativeness of the collected data in terms of scenarios and conditions encountered by 5G-capable users in 2020-2021 and 2023.

With regards to the Swedish part of the 2nd collection phase, we covered 7 IS locations (4 in Stockholm and 3 in Karlstad) and 3 OD locations (2 in Stockholm and 1 in Karlstad). Due to the focus on evolution aspects, only a portion of the data collected in Sweden was used for this paper, and particularly for the analysis of roaming in Section V.V-B.

2) PASSIVE NETWORK MONITORING

We used R&S TSMA6 to collect data for the analysis of RAN deployment, radio coverage, and positioning. We thus detected and decoded the DL signals broadcast by 4G LTE/LTE-A and 5G NR PCIs, by configuring R&S TSMA6 to monitor the following bands (reported from lower to higher frequency ranges): Band n28, 20, 32, 3, 1, 7, 42, and n78 (4G/5G common bands are indicated in LTE notation, i.e., without "n"). All bands use Frequency Division Duplex (FDD) except for Bands 42 and n78, which use Time Division Duplex (TDD). We detected four Italian MNOs with a 5G RAN deployed across both collection phases, which we refer to as Op_1^I, Op_2^I, Op_3^I, and Op_4^I (I is for Italy). During the 2nd phase, we detected two Swedish MNOs with significant 5G coverage in our measurements, which we refer to as Op_1^S and Op_2^S (S is for Sweden).

4G coverage measurements were executed on the Reference Signals (RSs) sent by 4G PCIs. Since 5G PCIs do not transmit RSs, a different approach was used for evaluating 5G coverage. The Physical Broadcast Channel (PBCH), the Primary Synchronization Signal (PSS), and the Secondary Synchronization Signal (SSS) are defined altogether as a Synchronization Signal Block (SSB), periodically sent by 5G PCIs within a bandwidth that depends on the Subcarrier Spacing (SCS) adopted in the underlying Orthogonal Frequency Division Multiplexing (OFDM) grid [30]. 5G PCIs can use SSB beamforming, so that different SSBs can be transmitted over narrow beams to increase spatial/user diversity and spectrum efficiency (up to 8 beams can be used in the mid band, as per Rel-15) [30].

R&S TSMA6 thus executed 5G coverage measurements at SSB beam level, hence providing a higher granularity compared to 4G.

3) ACTIVE PERFORMANCE TESTING

We used the UEs to execute tests on the MNOs' networks towards analyzing their QoS/QoE performance. In both collection phases, tests were repeated several times for each sub-campaign and location. In the following, we refer to a single test repetition as a *session*.

During the 1st collection phase in Italy, we focused on Op_1^I and Op_2^I as they were the only MNOs offering 5G services (Op_3^I and Op_4^I had 5G networks deployed, as highlighted by our passive network monitoring, but 5G commercial subscriptions were not publicly available yet). During the 2nd phase, we extended our tests to Op_3^I and Op_4^I, which in the meantime enabled 5G commercial use. In Sweden, we focused on Op_1^S and Op_2^S. For this paper we mostly analyze the Italian MNOs and use part of the measurements collected for the Swedish MNOs for dedicated analyses on roaming, which we performed by shipping SIM cards from Swedish MNOs to Italy, and vice versa.

With respect to connectivity, during both collection phases we alternatively configured our UEs to operate in two modes: i) *4G*: the UE only exposed its 4G capability so that MNOs could only connect it to 4G PCIs; and ii) *5G-enabled*: the UE exposed its 5G capability so that the MNOs could decide to connect it to a 5G PCI along with 4G PCIs.

1) Throughput Testing: During the 1st collection phase, we used Ookla Speedtest [31] to assess end-to-end DL and UL throughput. We configured the app to perform tests with multiple Transmission Control Protocol (TCP) connections and towards a server in the same city where tests were being executed. During the 2nd collection phase, we performed Ookla Speedtest along with another test compliant with the European Telecommunications Standards Institute (ETSI) specifications for throughput evaluation [32]. The ETSI test measures DL/UL throughput by downloading/uploading uncompressed files of 1 GB from/to a server in Switzerland, using several parallel Hypertext Transfer Protocol (HTTP) threads that saturate the channel capacity for 7 seconds [33].

2) Interactive Services Testing: We performed several tests for evaluating the achievable performance of beyond-eMBB interactive services. We used the Interactivity test in the R&S Qualipoc app, which employs a methodology approved by ETSI and International Telecommunication Union Telecommunication Standardization Sector (ITU-T) as the standard procedure for QoS/QoE testing on 5G systems [34], [35], [36].

The main idea is that the perceived responsiveness of real-time services is affected by three QoS metrics: latency, measured as Round Trip Time (RTT), Packet Delay Variation (PDV), i.e., jitter, and Packet Loss Rate (PLR), i.e., how many packets were lost or reached the destination after a given RTT budget during a test. A service-dependent QoE interactivity score (*i-score* [%]) can thus be defined as the perceived service responsiveness and evaluated as a function of these metrics. During a test, the UE sends a data stream to a server, which reflects packets back to the UE. The transport protocol used is User Datagram Protocol (UDP), while the higher layer protocol is an enhanced version of the Two-Way Active Measurement Protocol (TWAMP) [37], enabling reflection of packets of a different size compared to the ones received. Packet size and rate can thus be set and varied during a test to generate DL/UL traffic patterns emulating real services. RTT, PDV, and PLR are calculated on the packets, and the *i-score* can be evaluated using a model that accounts for service specificity. In particular, following [34], [35], the *i-score* model assumes that service responsiveness has a monotonous inverse dependency on RTT, with saturation areas at low and high RTT values. Therefore, a logistic function with service-specific parameters $f_{\max}$, a, and b is used to transform each RTT from non-lost packets into a value between 0% and 100%. Assuming N non-lost packets collected during a test, the RTT-dependent term of the *i-score* model, denoted $score_{\mathrm{RTT}}$, is evaluated as follows:

$$\begin{aligned} score_{\mathrm{RTT}} &= \frac{1}{N}\sum_{n=1}^{N}\frac{f_{\max}}{f_0}\left[1-\frac{1}{1+e^{-\frac{(\mathrm{RTT}_n-a)}{b}}}\right], \\ \text{where } f_0 &= 1-\frac{1}{1+e^{\frac{a}{b}}}. \end{aligned} \tag{1}$$

PDV and PER are then included via $score_{\mathrm{PDV}}$ and $score_{\mathrm{PER}}$ terms, which have service-specific parameters u and v, respectively, as follows:

$$\begin{cases} score_{\mathrm{PDV}} = \max\left(0, 1-\frac{\sigma_{\mathrm{PDV}}}{u}\right) \\ score_{\mathrm{PER}} = \max(0, 1-v\times \mathrm{PER}) \end{cases} \tag{2}$$

where σ_{PDV} is the standard deviation of the PDV evaluated on non-lost packets. Finally, *i-score* is defined as follows:

$$i\text{-}score = score_{\mathrm{RTT}} \times score_{\mathrm{PDV}} \times score_{\mathrm{PER}}. \tag{3}$$

During the 1st collection phase, we performed Interactivity tests towards a server in Switzerland and with a traffic pattern referred to as *eGaming real-time*. This is a DL/UL symmetric pattern that emulates phases of a typical online multi-player gaming application. Low-to-medium data rates (from 0.1 to 1 Mb/s) account for the fact that only status information is exchanged between UE and server, with video processing performed at the UE. Test duration is 10 seconds and the RTT budget to mark packets as lost is 100 ms, considering that 3GPP defines a maximum one-way delay of 50 ms for this application class in the 5G QoS Identifier (5QI) Class 3 [38]. During the 2nd collection phase, we also performed tests with a new pattern and using servers in Sweden. In particular, we used the *AR/VR Cloud Gaming* pattern. This is based on the traffic analysis of gaming platforms and emulates cloud services where the client sends status information to a server with a low bit rate (0.25 Mb/s in UL), the server processes high definition video (up to 1080p, 60 frames per second) and sends it back with high data rate (from 2 to 5 Mb/s in DL). Test duration and RTT budget are the same as for *eGaming real-time*. For both traffic patterns, we used $f_{\max} = 100$, $a = 61$, $b = 14$, $u = 120$, and $v = 4$, as specified in [35], [36].

3) Additional Tests: During both phases, we executed ping sessions towards either Google's public Domain Name System (DNS) resolver at 8.8.8.8 or the servers used for the previous tests, aiming to keep the UE radio connection active throughout the sub-campaigns. During the 2nd collection phase, we also performed traceroute tests to gain visibility on per-hop latency and identify potential network bottlenecks.

4) RAN CONFIGURATIONS AND RRC SIGNALING

The parallel use of ROMES in TSMA6 and Qualipoc in the UEs also allowed to monitor the RAN-UE interaction during our measurements. This is useful for in-depth analyses of the configurations adopted for specific network operations, including the management of 4G Carrier Aggregation (CA), 4G and 5G HOs, and 4G-5G DC.

Except for the information broadcast via PBCH, such configurations are exchanged via RRC signaling, which we thus decoded in both DL and UL directions. In particular, initial RAN information is transmitted in PBCH messages. These contain the Master Information Block (MIB), which informs UEs on basic settings and on where to find the first RRC message with more settings, i.e., the SIB1 (SIB stands for System Information Block). Besides providing cell-specific information (e.g., identifiers, selection configurations, TDD patterns, SCS, and SSB periodicity), SIB1 informs UEs on how to decode other SIBs, which provide further settings. It is worth highlighting that, differently from 4G and 5G SA, MNOs can avoid the transmission of SIBs by their 5G NSA PCIs. Since UEs are always connected to one 4G PCI acting as their Primary Cell (PCell), 5G RAN information can also be shared via RRC messages from the 4G PCell, along with other configurations, e.g., on how to measure channel quality and report HO measurements and events. Hence, in 5G NSA, it is key to monitor also RRC signaling, towards understanding how MNOs operate their networks. We used the decoding of RRC signaling to analyze the impact of HO and connectivity management on performance.

C. DATASET

The dataset collected by the TSMA6 during both collection phases includes spatial and temporal fields, frequency and cell identifiers (e.g., PCIs, and also SSB indexes for 5G), and signal strength and quality indicators, i.e., Reference Signal Received Power (RSRP) [dBm], Reference Signal Received Quality (RSRQ) [dB], and Signal to Interference and Noise Ratio (SINR) [dB]. These were measured on 4G RSs and different 5G control signals for all the PCIs/SSBs detected during each sub-campaign. Note that we used RSRP in our positioning analysis for defining the value of positioning *features*, as further described in Section VII.

TABLE 1. Number of passive samples collected in Rome during the 2nd collection phase (2023), for each technology, scenario, and MNO.

	4G		5G	
	IS	**OD**	**IS**	**OD**
Op_1^I	1.878M	632.8K	701.8K	211.9K
Op_2^I	1.896M	533.8K	4.99M	2M
Op_3^I	672.7K	334.9K	2.163M	795.7K
Op_4^I	983.7K	410.2K	3.751M	1.307M

The dataset collected by the UEs during both collection phases includes information on connection and coverage (e.g., RSRP, RSRQ, and SINR of the serving PCIs, i.e., the ones at which the UEs were connected), resource allocation (e.g., Modulation and Coding Scheme (MCS) and Transport Block Size (TBS)), and QoS/QoE performance (e.g., throughput at different layers for each throughput test session, and *i-score*, RTT, PDV, and PLR, for each Interactivity test session).

As mentioned above, the data collected during the 1st collection phase were already open-sourced in [4], where relevant statistics on the number of collected passive samples and executed active tests are also reported. In summary, we collected approximately 5.12M coverage samples for 5G (677K for 4G) across IS and OD locations, where each sample is a unique, geo-tagged point where the TSMA6 scanner was able to decode, at a given time, coverage information from the surrounding network deployments (one or more 4G/5G PCIs of one or more MNOs operating in one or more bands). Moreover, we executed 169/206 Ookla and 692/492 *eGaming real-time* sessions for $\text{Op}_1^I/\text{Op}_2^I$.

In conjunction with this paper, we open-source the dataset collected during the 2nd collection phase (see Note 1), and also provide a mapping for identifying the IS and OD locations that are in common across the two collection phases, aiming to ensure the replicability of our work as well as enable further comparative analyses. Table 1 summarizes the number of passive samples collected in Rome during the 2nd collection phase, for each technology (4G, 5G), scenario (IS, OD), and MNO (Op_1^I–Op_4^I). As reported in the table, we collected approximately 15.9M coverage samples for 5G (7.3M for 4G) across IS and OD locations. Moreover, Table 2 reports the number of test sessions executed in Rome during the 2nd collection phase, for each scenario and MNO, as a function of the adopted UE mode (*4G*, *5G-enabled*). For *eGaming real-time* and *AR/VR Cloud Gaming*, we report the number of sessions executed against the server in Switzerland. Finally, Table 3 reports the number of *eGaming real-time* and *AR/VR Cloud Gaming* sessions executed in Rome and Sweden during the 2nd collection phase, for each scenario and MNO, and with the UEs working in *5G-enabled* mode. In this case, we report the number of sessions executed against both servers in Switzerland and Sweden. When only considering the tests and MNOs also analyzed during the 1st collection phase, we executed 446/449 Ookla and 933/934 *eGaming real-time*

TABLE 2. Number of active sessions executed in Rome during the 2nd collection phase (2023), for each UE mode, MNO, test, and scenario. The total measurement duration for each test is provided alongside the test name (in hours). For *eGaming real-time* and *AR/VR Cloud Gaming*, we report the number of sessions executed against the server in Switzerland.

	4G				*5G-enabled*			
	Op_1^I	Op_2^I	Op_3^I	Op_4^I	Op_1^I	Op_2^I	Op_3^I	Op_4^I
ETSI Throughput Test (27 hr)								
IS	176	176	83	71	777	774	150	146
OD	77	76	17	17	167	179	66	78
eGaming real-time (26 hr)								
IS	111	110	95	95	768	770	85	83
OD	87	87	26	25	165	164	64	63
AR/VR Cloud Gaming (13 hr)								
IS	192	192	92	100	391	387	105	105
OD	41	42	N/A	N/A	127	127	70	70
Ookla (20 hr)								
IS	92	89	N/A	N/A	289	290	N/A	N/A
OD	17	20	N/A	N/A	48	50	N/A	N/A

(server in Switzerland) sessions for Op_1^I/Op_2^I. However, as clear from Tables 2 and 3, we intensified our measurement efforts during the 2nd collection phase, covering more tests and MNOs, and ultimately enabling our in-depth longitudinal analysis of 5G performance.

IV. NETWORK DEPLOYMENT AND THROUGHPUT PERFORMANCE

In this section, we study the changes over time of 5G NSA deployments and throughput performance. We carry out comparative analyses between the two collection phases, and also discuss additional insights derived from the larger sets of MNOs and tests covered during the 2nd collection phase. For simplicity, throughout this and the following sections, we refer to the 1st collection phase as "2021" and to the 2nd collection phase as "2023".

A. DEPLOYMENT

In the following, we first discuss RAN physical deployment and then explore the frequency perspective, aiming to reveal how MNOs leverage the frequency bands available for 5G services. Last, we provide insights on used channel bandwidth, SCS, TDD patterns, and SSB beamforming strategies.

1) RAN DEPLOYMENT

By using the R&S TSMA6 scanner, we identify the PCIs available at each of the locations where we performed measurements during both 2021 and 2023.

Table 4 shows the number of unique 4G and 5G PCIs detected across all sub-campaigns in IS and OD scenarios, for each MNO and collection phase (PCIs with samples constituting less than 1% of the data were ignored, since these PCIs were detected only sporadically and with very low signal strength). For this analysis, we focus on the frequency bands where PCIs were detected during both

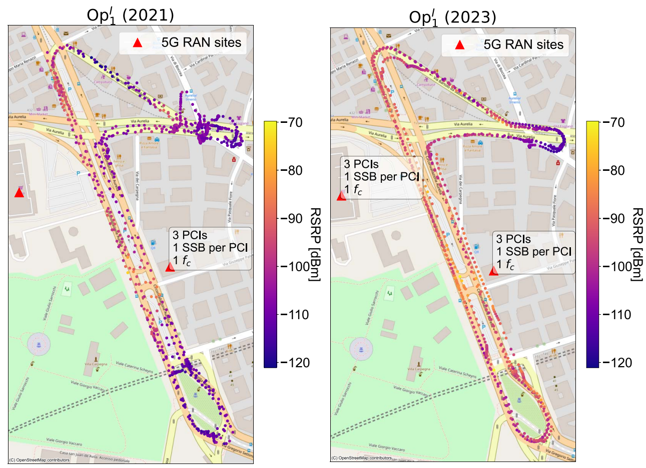


FIGURE 2. Spatial-temporal characterization of 5G RAN deployment and coverage for Op_1^I in the OD location, in 2021 (left) vs. 2023 (right). Coverage is reported as the highest RSRP value (across the PCIs detected for Op_1^I) observed in the points traversed during the measurements.

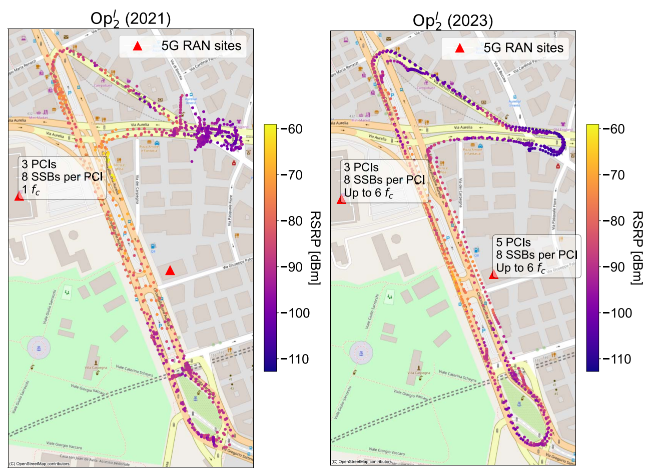


FIGURE 3. Spatial-temporal characterization of 5G RAN deployment and coverage for Op_2^I in the OD location, in 2021 (left) vs. 2023 (right). Coverage is reported as the highest RSRP value (across the PCIs detected for Op_2^I) observed in the points traversed during the measurements.

collection phases, i.e., n78 for 5G, and 3 and 7 for 4G. We observe that, in 2023, the average density across scenarios of deployed 5G PCIs for Op_1^I and Op_2^I has increased by 127% and 83%, respectively, while for Op_3^I and Op_4^I the respective increase is 38% and 60% (for Op_4^I, we only consider data from OD measurements since data from IS were available in 2021). Likewise, the average density of deployed 4G PCIs (across both scenarios) for Op_1^I and Op_2^I has increased by an average of 44% in 2023. On the contrary, there are no significant changes in the 4G RAN deployment for Op_3^I and Op_4^I. Our results show a clear trend towards denser 5G RAN deployments, as a result of ongoing investments of MNOs in providing services through 5G NSA deployments.

Fig. 2 and Fig. 3 provide a spatial-temporal characterization of the 5G RAN deployment of Op_1^I and Op_2^I observed from 2021 (left) to 2023 (right) in the OD location. Note

TABLE 3. Number of *eGaming real-time* and *AR/VR Cloud Gaming* sessions executed in Rome and Stockholm/Karlstad during the 2nd collection phase (2023), for each Italian and Swedish MNO, in roaming vs. not-roaming cases (UE mode is always *5G-enabled*). The total measurement duration for each test is provided alongside the test name (in hours). For both tests, we report the combined number of sessions executed against the servers in Switzerland and Sweden.

	Roaming						*Not-Roaming*					
	Op_1^I	Op_2^I	Op_3^I	Op_4^I	Op_1^S	Op_2^S	Op_1^I	Op_2^I	Op_3^I	Op_4^I	Op_1^S	Op_2^S
eGaming real-time (60 hr)												
IS	191	192	191	188	177	173	913	939	158	156	1092	1073
OD	N/A	N/A	N/A	N/A	133	130	285	294	132	132	216	230
AR/VR Cloud Gaming (46 hr)												
IS	196	203	197	194	217	210	811	837	192	191	918	902
OD	N/A	N/A	N/A	N/A	206	202	290	307	136	136	246	242

TABLE 4. Number of detected 4G and 5G PCIs for Op_1^I, Op_2^I, Op_3^I, and Op_4^I, grouped by scenario, technology, and collection phase (2021/2023). We highlight the highest value across the two collection phases using bold text.

	4G		5G	
	IS	**OD**	**IS**	**OD**
Op_1^I	12/**18**	11/**16**	4/**15**	7/**10**
Op_2^I	18/**24**	12/**18**	7/**11**	5/**11**
Op_3^I	**7**/6	10/10	5/5	3/**6**
Op_4^I	13/**16**	17/17	−/**13**	5/**8**

that we do not report figures for Op_3^I and Op_4^I, but the following analyses also hold for these two MNOs. With respect to Op_1^I, out of the 7 PCIs detected by our scanner in 2021 (Table 4), we are able to precisely locate 3 PCIs, all deployed on the same RAN site (also referred to as *tower*). In 2023, these PCIs are complemented by new PCIs, forming an overall set of 10 PCIs (Table 4), with 3 new PCIs active on a different tower. Similar observations hold for Op_2^I, with slightly different numbers. In particular, we observe that Op_2^I has 5 new PCIs active on a new tower in 2023. A joint analysis between MNOs further shows that, around the OD location, Op_1^I and Op_2^I use tower sharing, with the two identified towers used by both MNOs to deploy their PCIs.

Fig. 2 and Fig. 3 also provide an indication of the 5G radio coverage observed as a result of the RAN deployments described above, in terms of the highest RSRP measured in each point forming the OD location. As further quantified later while discussing throughput results, we observe clear coverage gains in 2023, with higher RSRP levels measured in several portions of the OD location, for both MNOs.

2) FREQUENCY DEPLOYMENT

Next, we compare the distribution of data over the different frequency bands. Our data show that, in the mid band (i.e., Band n78), Op_1^I and Op_2^I use a bandwidth of 80 MHz, Op_3^I of 20 MHz, and Op_4^I of 60 MHz, all with a SCS of 30 kHz. Moreover, in 2023, some MNOs transmit their SSB signals on multiple carrier frequencies (f_c) within the same band, compared to a single f_c found in 2021. As a reference example, we report in Fig. 2 that a single f_c is used by Op_1^I for transmitting SSBs around our OD location, during both 2021 and 2023; on the other hand, as reported in Fig. 3, Op_2^I has increased the number of f_c, with up to 6 frequencies used to transmit different SSBs in 2023.

Unlike 2021, 1% of the data for Op_1^I and Op_2^I, and 37% of the data for Op_3^I were recorded in the low band (Band n28) in 2023. Similarly, 36% of Op_4^I data was recorded in Bands 3, 1, and 7 in 2023. Indeed, we further verified that Op_4^I leverages additional 20 MHz in the low band, shared with 4G. Finally, compared to 2021, the scanner identified 4G data for Op_1^I and Op_2^I in Band 32 during 2023. Overall, our results indicate that, in 2023, MNOs leverage more bands and frequency carriers to meet the demands from their users.

3) OTHER DEPLOYMENT CONFIGURATIONS

Last, we discuss other deployment configurations adopted by MNOs. In the time domain, we find that all MNOs use the same TDD pattern, *DDDDDDDFUU*, where *D* represents a DL slot, *U* an UL slot, and *F* is a flexible slot used in both DL/UL (same for all MNOs, 6 DL symbols and 4 UL symbols). We also observe that Op_1^I does not support SSB beamforming, Op_2^I and Op_3^I support SSB beamforming with up to 8 beams, while Op_4^I supports it with up to 6 beams. Across the two collection phases, no significant changes on the above strategies were observed, as also exemplified in Fig. 2 and 3 for the OD location, where we report that Op_1^I and Op_2^I did not change their beamforming strategies from 2021 to 2023. As discussed next, the MNO configuration differences provide valuable insights towards interpreting the performance disparity observed among MNOs.

B. THROUGHPUT

We now focus on assessing the throughput performance observed in 5G NSA deployments. Using Ookla Speedtest, we first study disparities between Op_1^I and Op_2^I in 2021 and 2023. Then, we reveal further insights by using a more controlled ETSI-compliant throughput test over a larger pool of Italian MNOs covered in 2023.

1) THROUGHPUT PERFORMANCE OVER TIME

Fig. 4 shows DL/UL throughput performance (based on Ookla Speedtest) for Op_1^I and Op_2^I, per scenario and collection phase. We mainly analyze the performance in *5G-enabled* tests, with results for *4G* tests in 2023 reported for further comparison.

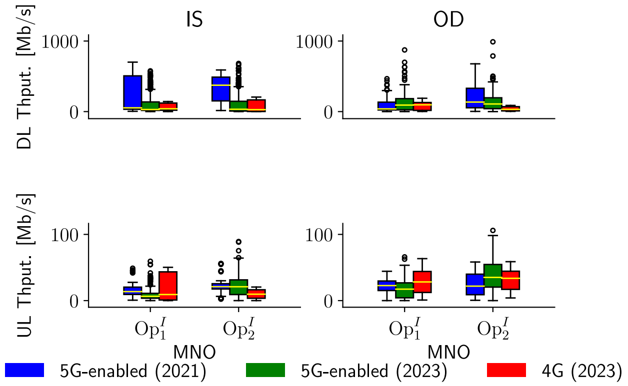


FIGURE 4. DL (top) and UL (bottom) throughput performance obtained during *5G-enabled* Ookla Speedtests, for Op_1^I and Op_2^I in 2021 vs. 2023, in IS (left) and OD (right) scenarios. Performance for *4G* tests in 2023 is reported for further comparison.

For IS, we observe a DL throughput decrease for both MNOs in 2023, particularly evident for Op_2^I. In the UL case, Op_1^I experiences a visible throughput decrease in 2023, while Op_2^I has a median performance only slightly lower in 2023 compared to 2021. For OD, results are more varied across MNOs and scenarios: Op_1^I shows DL throughput gains in 2023 compared to 2021, but losses in UL throughput; the opposite behaviour is observed for Op_2^I. Comparing with 4G, similar general performance is observed in 2023 between *5G-enabled* and *4G* tests, but Op_2^I seems to benefit more from 5G in UL (in IS and OD) and DL (in OD). Overall, 5G results predominantly show performance losses when 2023 is compared to 2021 (e.g., in DL for both MNOs in IS), but also some gains (e.g., in UL for Op_2^I in OD).

We perform Kruskal-Wallis and Dunn's tests to determine the statistical significance of the difference between *5G-enabled* throughput results in 2021 and 2023. On the one hand, the difference between 2021 and 2023 is statistically significant for both MNOs and DL/UL scenarios in IS. On the other hand, no statistical significance is obtained in OD, which is alignment with the more varied results reported in Fig. 4 across MNOs and DL/UL cases, and may also depend on the lower amount of data available for this scenario.

In order to isolate key factors for the observed performance disparity, we perform in-depth analyses of our measurements. We identify two root causes that contribute to the 5G performance decrease in 2023, as well as one main cause resulting in the UL gains for Op_2^I. We start with the causes for the performance loss and introduce them via two examples. In Fig. 5, we compare the time series collected for DL throughput (top row) and further metrics (other rows) during one *5G-enabled* sub-campaign in 2021 (left column) and two sub-campaigns in 2023 (middle/right columns), performed in the same IS location for Op_1^I. The additional metrics include MCS, TBS, RSRP, and SINR, for both 4G and 5G. Indeed, thanks to DC, both 4G and 5G can be used simultaneously, resulting in contributions from 4G and 5G Physical Downlink Shared Channel (PDSCH) that sum up to the application throughput.

In 2021, we observe a DL throughput consistently exceeding 500 Mb/s in all test sessions whereas, in both 2023 sub-campaigns, the DL throughput always remains below 500 Mb/s. MCS and TBS also show higher values in 2021 compared to 2023 (for TBS, we refer to the 5G values, as data for 4G were not available in 2021). The cause is shown in the bottom figures, where we see that 5G RSRP and both 4G and 5G SINR are lower by several dBs in both 2023 sub-campaigns compared to 2021, thus resulting in lower allocated MCS values and, in turn, lower TBS and throughput. Op_2^I shows similar results, therefore, we conclude that the first verified cause for DL throughput performance losses is the worsening of the coverage conditions.

Fig. 6(a) validates the observed decline in coverage for the IS scenario in 2023. In particular, we observe that the median 5G RSRP for both Op_1^I and Op_2^I in IS scenarios was higher in 2021 compared to 2023. This result includes all PCIs at which our UEs were connected during all sub-campaigns of our collection phases. Similar trends are observed for 5G SINR (Fig. 6(b)). To determine the statistical significance of the results in Figure 6, we leverage the Wilcoxon test. We thus perform pairwise comparisons between scenarios, MNOs, and collection phases. The resulting p-values reveal that a statistically significant difference exists in all cases.

Focusing on the second sub-campaign in 2023 (right column in Fig. 5), we see an even more pronounced throughput reduction after two test sessions. As a matter of fact, in this test, 5G connectivity is completely lost after these two sessions, with 5G PDSCH no longer contributing to the overall throughput, leading to a clear negative effect due to the full fallback on 4G. By further inspecting our measurements, we verify that this 5G connectivity issue (as well as others observed during our tests) happens when the UE informs the network of a Secondary Cell Group (SCG) Failure and, as a result, gets disconnected from the 5G RAN. Hence, the second verified root cause for DL throughput performance losses is the presence of SCG Failures. We provide more details on this aspect in Section VI, where we first illustrate that this issue is also related to MNOs' HO configuration policies, and then quantify its negative effect not only on throughput performance but also on the QoS/QoE of interactive services.

Next, we examine the UL throughput gains of Op_2^I, and show that a better UL usage of 4G-5G DC is the cause for the 2023 improvement. Fig. 7 compares the UL use of DC of Op_2^I in 2021 (a) and 2023 (b), by reporting the experienced UL throughput along with the contributions from 4G and 5G Physical Uplink Shared Channel (PUSCH). Fig. 7(a) shows that, in 2021, Op_2^I exclusively used 5G PUSCH when connected to 5G. This is different in 2023, as shown in Fig. 7(b), with Op_2^I predominantly leveraging 4G PUSCH while also exploiting 5G PUSCH. Hence, possible UL coverage limitations for 5G are better handled in 2023, where DC with a combined use of 4G and 5G resources positively impacts performance and leads

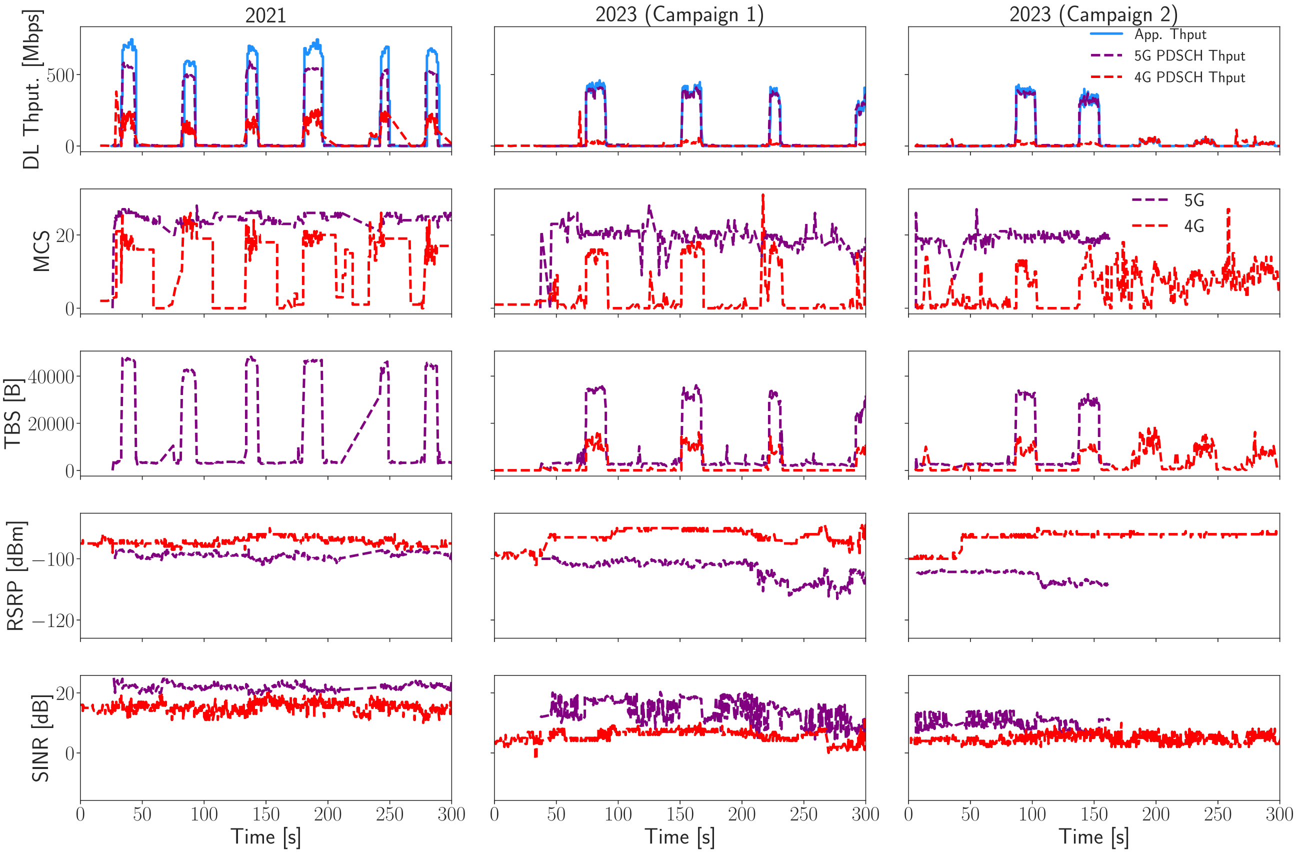


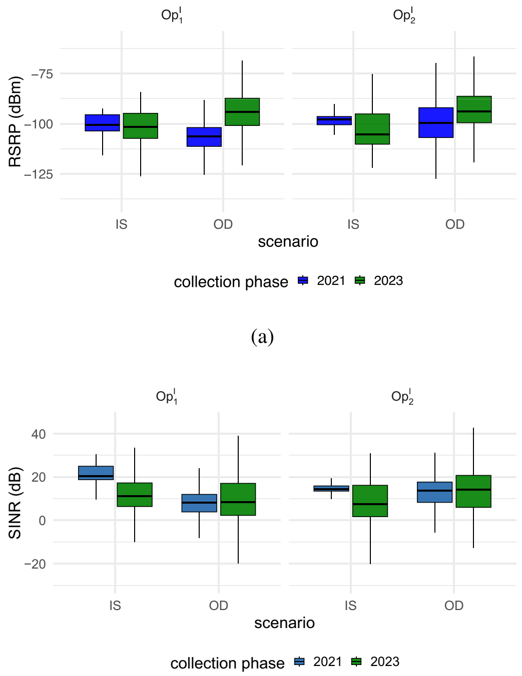
FIGURE 5. Time series of DL throughput, MCS, TBS, RSRP, and SINR during one *5G-enabled* sub-campaign in 2021 (left column) and two sub-campaigns in 2023 (middle/right columns) for Op^I_1 in the same IS location (Data for 4G TBS not available for this campaign in 2021).

to a higher UL throughput for Op^I_2, as also reported in Fig. 4.

As an additional analysis, we verify the existing linear correlation between 5G PDSCH/PUSCH throughput, network-allocated resources (TBS and MCS), and observed coverage conditions (RSRP and SINR). For this analysis, we focus on 2023 but similar observations can be derived for 2021. Fig. 8 (resp. Fig. 9) shows scatter plots between the median MCS, TBS, RSRP, SINR, and 5G PDSCH (resp. 5G PUSCH) throughput observed during each Ookla and ETSI-compliant throughput test session, for both MNOs and IS/OD scenarios. The Pearson's correlation value and its statistical significance (p-value) are also reported, together with the best linear fit for the observed data. Results show that both DL and UL throughput linearly depend on the allocated MCS and, thus, on the resulting TBS, with correlation values always above 0.60. Considering coverage, we observe a linear correlation value of about 0.60 for RSRP in both DL/UL, which decreases to about 0.40 (DL) and 0.20 (UL) for SINR. This result highlights that, as expected, coverage has a significant impact on throughput performance; non-linear interdependencies, however, exist between RSRP/SINR and throughput, due to multiple non-linear factors affecting signal propagation and, in turn, throughput, which also aligns with previous empirical observations (e.g., [26]). In both DL/UL, RSRP is still more linearly correlated than SINR, which also reflects the fact that, as we verified on both MNOs' networks, the majority of operations (e.g., HO and connectivity management) are RSRP-based rather than SINR-based (in fact, 3GPP does not specify methodologies for calculating the SINR, which is left to UE manufacturers), ultimately leading to higher linear correlations between RSRP and throughput.

2) THE ROLE OF 4G-5G DC ON THROUGHPUT PERFORMANCE

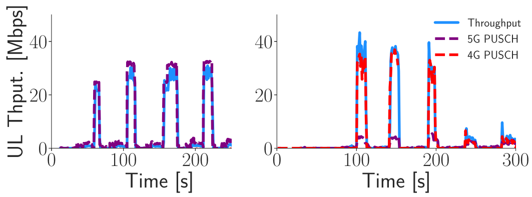
In the previous section, by using *5G-enabled* tests we show that, depending on coverage conditions, 5G connectivity (e.g., SCG Failures), and 4G-5G DC strategies, data may be completely or only partially (or not at all) exchanged with a 5G PCI during a test session. To gain further insights on the role of 4G-5G DC on performance, we analyze this aspect more in-depth. We label test sessions based on the utilization level of 5G PDSCH (PUSCH for UL), eventually labeling sessions as *5G Partial* if 5G PDSCH/PUSCH was used for less than 80% of its duration, and as *5G* for a utilization higher than 80%. We perform this analysis on the ETSI-compliant throughput tests, considering that these are more controlled tests compared to Ookla Speedtests, with clear DL/UL time separation and fixed duration. Moreover, we use 2023 measurements to also analyze Op^I_3 and Op^I_4.



(a)

(b)

FIGURE 6. RSRP (a) and SINR (b) across all serving 5G PCIs, grouped by MNO, scenario, and collection phase.



(a) (b)

FIGURE 7. Effect on UL throughput of different 4G-5G DC strategies of Op_2^I, in 2021 (a) and 2023 (b).

Fig. 10 shows DL (top) and UL (bottom) throughput performance for all Italian MNOs in IS (left) and OD (right) scenarios, with sessions labeled by using the methodology described above. We consistently observe higher throughput during *5G* sessions compared to *5G Partial*, which further highlights the benefit of fully exploiting 5G PDSCH/PUSCH resources and, in turn, the drawbacks from unstable 5G connectivity, e.g., due to 4G fallback after SCG Failures. This is in line with previous results in this paper and in other studies (e.g., [6], [8], [9]). We also observe better throughput in the OD scenario compared to the IS case, as an effect of better coverage conditions (Fig. 6). Fig. 10 also shows significant lower performance for Op_3^I and Op_4^I compared to Op_1^I and Op_2^I, which is well explained by the lower amount of frequency resources owned by these two MNOs, as reported in our previous analysis on deployment.

V. INTERACTIVE SERVICES AND ROAMING PERFORMANCE

In this section, we study 5G NSA networks from interactive services and roaming perspectives. Similar to the previous section, we first compare performance across 2021 and 2023, and then discuss new insights derived on more MNOs and by using the additional tests executed in 2023.

A. INTERACTIVE SERVICES

In the following, we assess QoS/QoE performance observed for interactive services on 5G NSA networks. Using the *eGaming real-time* traffic pattern, we first analyze performance disparities between Op_1^I and Op_2^I in 2021 and 2023, and then reveal more insights by exploring the 2023 measurements for all Italian MNOs. Finally, we extend our understanding on interactive services by analyzing the results obtained in 2023 by executing interactivity tests with the *AR/VR Cloud Gaming* traffic pattern.

1) eGAMING REAL-TIME PERFORMANCE OVER TIME

Fig. 11 shows QoS/QoE performance observed during *eGaming real-time* tests for Op_1^I and Op_2^I, per scenario and collection phase, and with the server in Switzerland. Similar to the throughput case, we analyze the performance in *5G-enabled* tests, with results for *4G* tests in 2023 reported for comparison. We show the per-session median results for three QoS metrics, i.e., RTT (top left), PDV (top right), and PLR (bottom left), and for the QoE metric evaluated on top of them, i.e., the *i-score* (bottom right).

Fig. 11 shows a significant *i-score* decrease in 2023 for both MNOs, as a direct result of higher RTT and PDV values. The only improvement we observe is for Op_2^I in the OD scenario, with a higher *i-score* due to the lower PLR observed in 2023, compared to rather large values experienced in 2021. Note that the same insights discussed on coverage in the throughput sections also apply to these tests, considering that Fig. 6 shows agglomerated coverage results, across all sub-campaigns and independent on tests being executed on the UEs. Additionally, SCG Failures also affected the *eGaming real-time* tests negatively, although to a lesser extent, as detailed in the analysis reported in Section VI.

We perform Kruskal-Wallis and Dunn's tests to determine the statistical significance of the difference between *5G-enabled i-score* results. In the IS case, the differences between 2021 and 2023 are statistically significant for both MNOs. In the OD case, the statistical significance is verified for Op_1^I but not for Op_2^I, which indeed shows the only case of *i-score* increase from 2021 to 2023.

We also verify the existing linear correlation between *i-score*, RTT and PLR QoS metrics, and observed coverage conditions (RSRP and SINR). We focus on 2023 but similar observations can be derived for 2021. Moreover, we leave out the PDV metric since we find PDV results in line with the ones observed for RTT. Fig. 12 shows scatter plots between the median RTT, PLR, SINR, and RSRP, and the *i-score* observed during each *5G-enabled eGaming*

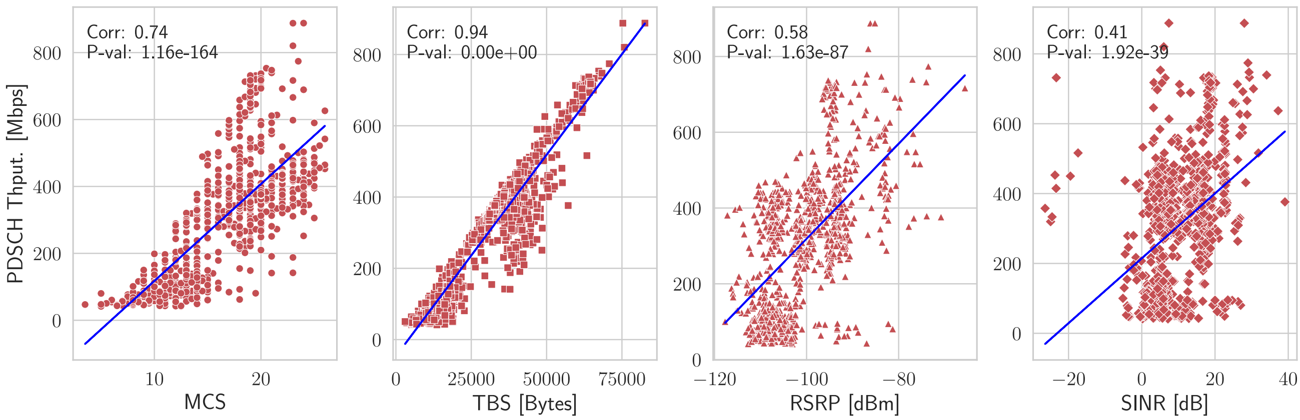


FIGURE 8. Scatter plots between the median MCS, TBS, RSRP, SINR and 5G PDSCH (DL) throughput observed during each Ookla and ETSI-compliant throughput test session in 2023, for Op_1^I and Op_2^I and IS/OD scenarios (MCS, TBS, RSRP, and SINR from left to right). The Pearson's correlation value and its statistical significance (p-value) are also reported, together with the best linear fit for the observed data (blue line).

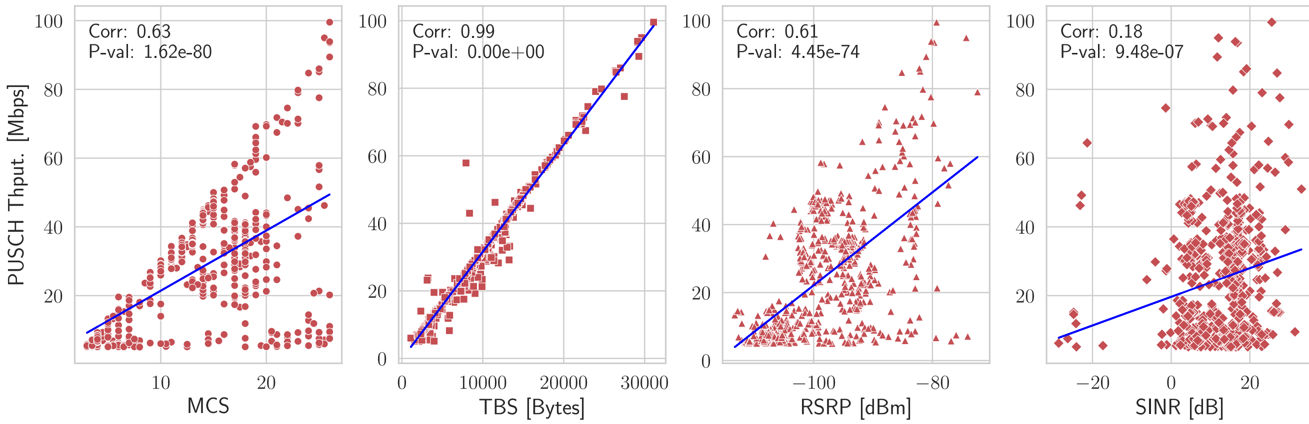


FIGURE 9. Scatter plots between the median MCS, TBS, RSRP, SINR and 5G PUSCH (UL) throughput observed during each Ookla and ETSI-compliant throughput test session in 2023, for Op_1^I and Op_2^I and IS/OD scenarios (MCS, TBS, RSRP, and SINR from left to right). The Pearson's correlation value and its statistical significance (p-value) are also reported, together with the best linear fit for the observed data (blue line).

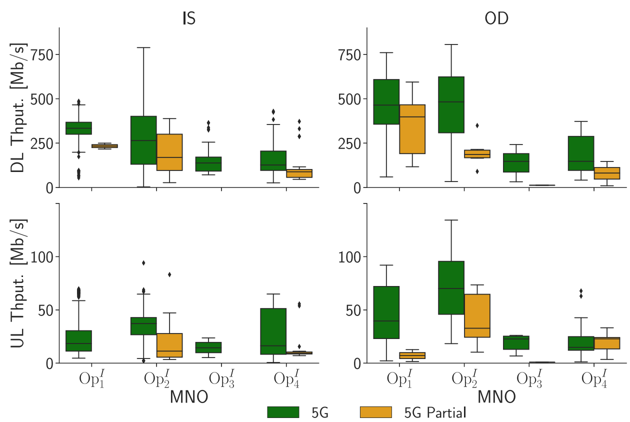


FIGURE 10. DL (top) and UL (bottom) throughput for *5G* and *5G Partial* sessions for all Italian MNOs in 2023, in IS (left) and OD (right) scenarios.

real-time session, for Op_1^I and Op_2^I, IS/OD scenarios, and the server in Switzerland. The Pearson's correlation value and its statistical significance (p-value) are also reported, together with the best linear fit for the observed data. Results show a clear inverse dependency between *i-score* and RTT, as a direct result of the $score_{RTT}$ definition in (1). Similarly, due to the $score_{PLR}$ definition in (2), PLR does not impact *i-score* linearly but a clear opposite trend is observed when PLR is between 0 and 15% (*i-score* is always 0% when PLR $>$ 15%). Considering coverage, we observe relatively low linear correlation values between *i-score* and RSRP/SINR. Similar to the throughput case, this result highlights that, although coverage impacts performance, non-linear interdependencies exist between RSRP/SINR, *i-score*, and the underlying QoS metrics (e.g., RTT and PLR), which are not reflected in the linear correlation coefficient. More in-depth investigations are needed on this aspect, which are made possible across the research community thanks to our open-sourced dataset.

2) THE ROLE OF 4G-5G DC ON eGAMING REAL-TIME PERFORMANCE

We now further explore *eGaming real-time* performance towards better analyzing the role of 4G-5G DC. Hence,

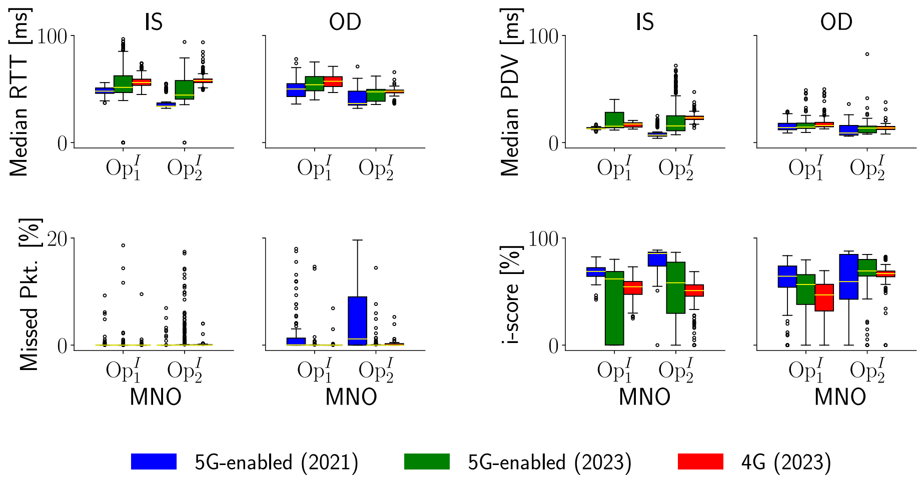


FIGURE 11. QoS/QoE performance obtained during *5G-enabled eGaming real-time* tests, for Op_1^I and Op_2^I in 2021 vs. 2023 (server in Switzerland). Per-session median RTT (top left), median PDV (top right), PLR (referred to as Missed Pkt., bottom left), and *i-score* (bottom right). For each metric, performance in IS (left) and OD (right) scenarios is reported. In all cases, performance for *4G* tests in 2023 is reported for further comparison.

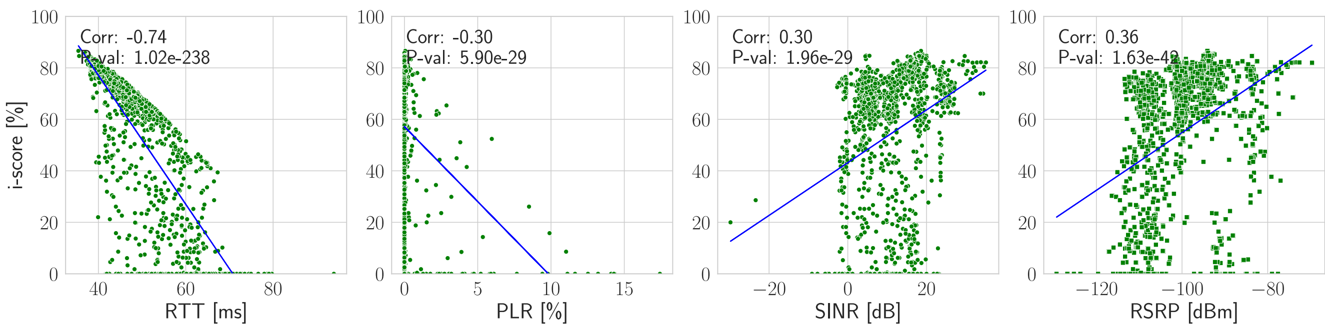


FIGURE 12. Scatter plots between the median RTT, PLR, SINR, and RSRP, and the *i-score* observed during each *5G-enabled eGaming real-time* session in 2023, for Op_1^I and Op_2^I, IS/OD scenarios, and the server in Switzerland (RTT, PLR, SINR, and RSRP from left to right). The Pearson's correlation value and its statistical significance (p-value) are also reported, together with the best linear fit for the observed data (blue line).

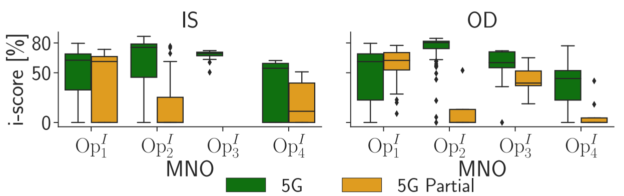


FIGURE 13. *eGaming real-time i-score* for *5G* and *5G Partial* sessions for all Italian MNOs in 2023, in IS (left) and OD (right) scenarios (server in Switzerland).

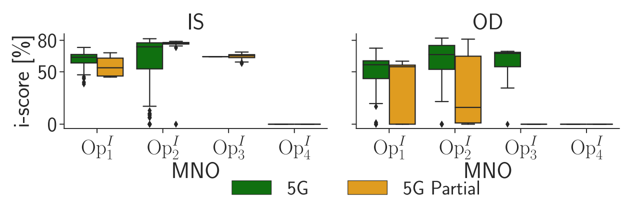


FIGURE 14. *AR/VR Cloud Gaming i-score* for *5G* and *5G Partial* sessions for all Italian MNOs in 2023, in IS (left) and OD (right) scenarios (server in Switzerland).

similar to the throughput case, we split sessions in *5G* and *5G Partial*, and also consider Op_3^I and Op_4^I. Fig. 13 shows the *i-score* performance observed during *eGaming real-time 5G* vs. *5G Partial* sessions, for all Italian MNOs in IS (left) and OD (right) scenarios (server in Switzerland). Better *i-score* performance is consistently observed during *5G* sessions, for all MNOs and scenarios (apart from the Op_1^I / OD case, where *5G* and *5G Partial* have similar median performance). On the one hand, Op_2^I achieves the highest *i-score*, with a median value reaching 80% in both IS and OD scenarios, as a result of the lowest median RTT observed in those cases. On the other hand, we observe significantly lower performance for Op_4^I, as a result of a highest median RTT observed during the tests for this MNO.

3) AR/VR CLOUD GAMING SERVICES PERFORMANCE

We next extend our analysis of interactive services to the more DL data-intensive case of AR/VR. Fig. 14 shows the *i-score* of *AR/VR Cloud Gaming* tests across the Italian MNOs, in IS and OD scenario and using the server in Switzerland. Compared to *eGaming real-time*, performance is slightly poorer across all MNOs, with Op_4^I now failing completely (*i-score* $= 0\%$) in both scenarios. Despite the latency budget being identical for both real-time gaming and AR/VR applications (100 ms), the latter demands substantially higher data rates due to the necessity of transmitting high definition video. This requirement particularly contributes to the poor *i-score* observed for Op_4^I. Our analysis

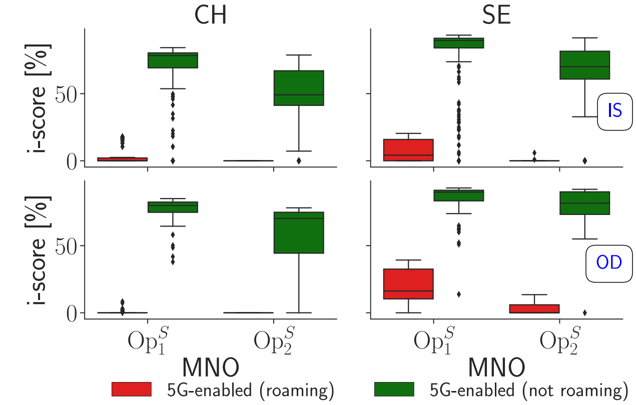


FIGURE 15. ***eGaming real-time i-score*** **for the Swedish MNOs (IS/OD scenarios) in roaming vs. not roaming cases. Results per server location (CH, SE) are in different columns.**

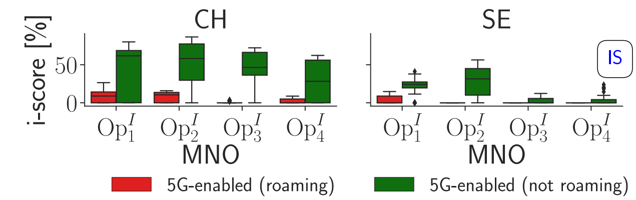


FIGURE 16. ***eGaming real-time i-score*** **for the Italian MNOs (IS scenario) in roaming vs. not roaming cases. Results per server location (CH, SE) are in different columns.**

shows a very high PLR, with median values exceeding 50% in the case of Op_4^I (recall that packets delayed beyond the latency budget also contribute to the PLR), highlighting the challenges faced in maintaining satisfactory QoS/QoE for these services in current 5G deployments.

As a further test, we conduct measurements with another traffic pattern emulating industrial process automation services with low data rates but with very strict latency requirements (20 ms) [39]. None of the tests yield positive outcomes (*i-score* always 0%), thus confirming that such URLLC services are not supported in current commercial 5G networks.

B. ROAMING

Last, we discuss the performance of roaming services in 5G NSA networks in 2023, by quantifying the roaming effect on interactive services. In particular, we evaluate the roaming performance across MNOs with 5G-enabled UEs, by examining the performance of Swedish MNOs in Italy and Italian MNOs in Sweden.

Our measurements reveal that, among Italian MNOs, Op_1^I and Op_3^I provide 5G roaming support, along with Swedish MNO Op_1^S. Conversely, Op_2^I and Op_4^I from Italy and Op_2^S from Sweden do not offer 5G roaming and fall back to 4G during roaming. Moreover, our observations indicate that all the MNOs use home-routed roaming, similar to what is observed in 4G networks [40].

Fig. 15 and Fig. 16 show the roaming performance measured via *i-score* during real-time gaming tests, for Italian and Swedish MNOs across IS and OD scenarios, in comparison to non-roaming situations. For this analysis, we use servers located in Switzerland (CH) and Sweden (SE), to gauge the impact of server proximity and roaming on service performance. Our results highlight a significant performance decrease due to roaming, as a consequence of the increased RTT resulting from routing data via MNOs' home networks. Fig. 15 shows that when the Swedish MNOs roam in Italy, performance significantly suffers with median *i-score* dropping to 0% in many instances, as opposed to performance within Sweden. We also observe that using the SE server helps but it is still insufficient, with the best roaming performance for Op_1^S seen in the OD scenario achieving a median *i-score* of 16%. The figure also highlights that a server close to the home network (SE vs. CH) offers consistent advantages, resulting in a 10% increase in median *i-score* for the non-roaming cases. Similarly, Fig. 16 shows significant median *i-score* decreases (0% in many cases) for the Italian MNOs while roaming in Sweden, compared to performance within Italy. Nevertheless, having a server closer to the home network, CH in this case, also provides some improvements, notably reducing RTT by up to 30 ms for Op_3^I and Op_4^I, as confirmed by our traceroute analysis. Yet, even with this benefit, Op_1^I only manages to achieve a median *i-score* of about 9%, the highest among all roaming instances. Fig. 16 also shows that a close server to home network (CH vs. SE) yields consistent benefits, leading to about 36% increase in median *i-score* in the non-roaming cases. We observed a similar behaviour when comparing the performance of roaming vs. not roaming for AR/VR services.

VI. HO AND CONNECTIVITY MANAGEMENT

We now provide a focused analysis on HO and connectivity management in the 5G NSA networks of Op_1^I and Op_2^I, highlighting how changes in such functions contributed to the performance losses discussed in previous sections.

In 5G NSA deployments, HO and connectivity management is challenging due to the hybrid nature of the RAN, where DC between 4G and 5G PCIs (also referred to as *cells* in this section, to better map with common terminology) can potentially bring significant performance improvements if carefully configured. Within 5G NSA networks, UEs use *cellsets* [25] formed by i) one 4G PCell, managing cellset composition and changes, ii) one or more 4G Secondary Cells (SCells), enabling CA, and iii) one or more 5G cells, enabling DC and forming the SCG (current NSA deployments predominantly use a single 5G cell, as also verified in our measurements in both 2021 and 2023). Following the RRC configurations shared by the PCell, UEs perform HOs and change cellsets, going through three main steps [19]:

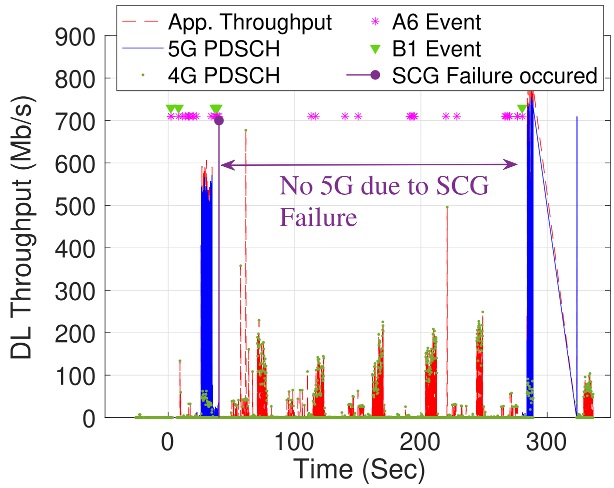
1) *Configuration:* UEs receive instructions from the PCell on the parameters and conditions to use for measuring and reporting HO-relevant information.

2) *Measurement and Reporting:* UEs measure relevant information and provide it to the PCell as measurement reports, commonly referred to as *HO events*.
3) *Decision and Execution:* The PCell uses HO events to decide if HOs are needed. Decisions are sent to the UEs.

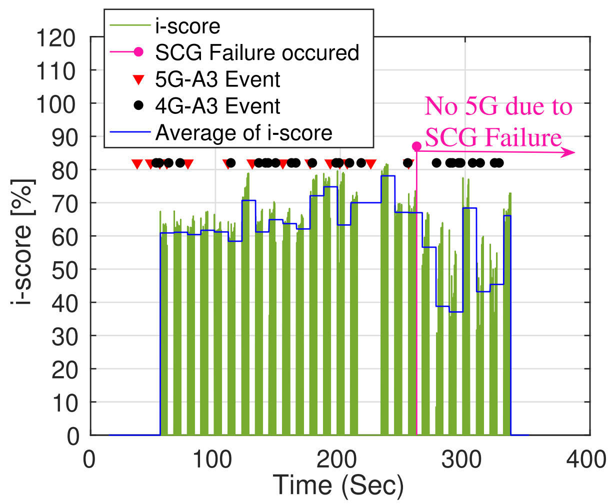
The definition of HO events in 5G NSA follows LTE standards [20]. Among several HO events that networks can configure and use, our combined analysis of RRC configuration messages, HO events, and HO decisions shows four key HO events ruling the connectivity in the 5G NSA networks of Op_1^I and Op_2^I, both in 2021 and 2023: i) *A*3, triggering changes of 4G PCells (4G→4G) and of 5G cells (5G→5G), ii) *A*6 and *A*4, triggering changes in the set of 4G SCells (4G→4G), and iii) *B*1, triggering inter-technology HOs, i.e., 5G cell additions (4G→5G).

By analyzing the configurations for the above events in 2021 and 2023, we find significant changes for 5G-related HOs, i.e., A3 and B1. On the one hand, Op_1^I did not have 5G A3 events configured in 2021, meaning that direct 5G cell changes were not possible; in 2023, it instead has these events configured, with a report being triggered if the UE finds a 5G neighbor cell having an RSRP 7 dB higher than that of the current 5G cell. On the other hand, Op_2^I changed 5G A3 configurations, with events being triggered with a 3 dB lower threshold in 2023. We also find changes in the B1 event for both MNOs, with lower RSRP thresholds used in 2023 compared to 2021. These changes suggest a more dynamic and aggressive use of the 5G RAN, also due to the lower thresholds triggering B1 events. As a result, we observe that, in 2023, PCIs with lower RSRPs are being used as 5G cells, thus demonstrating the willingness of MNOs to let 5G UEs use 5G RAN in worse coverage conditions.

We argue that this decision could have contributed to the more marked presence of SCG Failures in 2023 compared to 2021, with the negative effects on user performance preliminary observed above and further quantified below. When SCG Failures occur, the 5G cell connection is suspended and the UE reports SCG Failure information to the PCell, which in turn decides whether to release, reconfigure, or change the 5G cell. In our measurements, we mostly observe 5G cells being released after SCG Failures. Among many possible standard reasons for SCG Failures, we always observe *rlc-MaxNumRetx* as the reason, which hints at UL transmission issues potentially due to bad coverage, with the UE trying to send data to the 5G cell but reaching the maximum number of allowed retransmissions [41]. Compared to 2021, in 2023 we also observe more often that, after SCG Failures, the PCell reacts not only by releasing the 5G cell but also canceling the configurations for B1 events, blocking in this way a new addition of the same or of a different cell. Different timers are used by different PCells in this process, which potentially hinders 5G connectivity for a long time.



(a)



(b)

FIGURE 17. Examples of the impact of SCG Failures on (a) DL throughput and (b) *i-score*.

Fig. 17 shows two examples of the impact of SCG Failures on (a) DL throughput (Ookla Speedtest) and (b) *i-score* (*eGaming real-time* tests), observed for Op_1^I. On the one hand, the release of a 5G cell due to SCG Failure significantly affects DL throughput, because the additional bandwidth provided by 5G cannot be exploited. On the other hand, the *i-score* also drops after the SCG Failure, due to the higher latency observed over 4G, on which the UE falls back after the SCG Failure. By averaging over all occurrences of SCG Failures, we quantify a 60% throughput loss (both MNOs) and a 20% *i-score* loss (only Op_1^I, no SCG Failures were observed for Op_2^I during *eGaming real-time* tests).

VII. POSITIONING

The ability of 5G to provide precise location information is crucial for a variety of services, including navigation, monitoring, and tracking.

While 5G positioning was introduced in 3GPP Rel-16, with additional features introduced in subsequent releases aimed at improving positioning performance, most commercial networks are still based on Rel-15, lacking positioning support at the network side. In this context, we utilized the collected 5G coverage data to implement a positioning technique that can operate with data collected at the UE side only; the technique is based on fingerprinting [42]. Fingerprinting takes its name from the idea of associating to each location a fingerprint, defined as the values recorded for a set of *features* related to radio coverage; in general, only a subset of the features will be detected at any given location, and the fingerprint will be completed by adopting a default value for missing features. Fingerprinting-based positioning operates in two phases: the *offline phase*, where a database of fingerprints is created by collecting data at reference locations, usually referred to as Reference Points (RPs), and the *online* phase, where the fingerprint provided by a target device is used to estimate its position. Estimation can be done, e.g., via the Weighted k Nearest Neighbor method, where the target fingerprint is compared against the RP database to determine the $k \geq 1$ better matching fingerprints, and the position of the device is obtained as a weighted mean of the corresponding k RPs.

Within our framework, a feature is defined as a combination, unique throughout the collected dataset, of the following identifiers: PCI, SSB Index, New Radio Absolute Frequency Channel Number (NRAFCN), which corresponds to the SSB carrier frequency f_c defined in Section IV-A, and Mobile Network Code (MNC). Data for the same OD location were used for 2021 vs. 2023, using a nearly identical number of measurement points (1192 for 2021 vs. 1193 for 2023), that were divided in RPs and Test Points (TPs). The performance indicator used in the analysis was the Minimum Average Positioning Error, defined as the minimum of the average positioning error over all the TPs as a function of k.

For each feature, we used the RSRP measured on the 5G SSS collected as part of the coverage data, to define its value. We carried out our analysis focusing on the impact of *feature density* (number of features that define a fingerprint) and *spatial density* (number of RPs in the positioning service area). Both densities are expected to improve positioning accuracy as they increase, but while the latter is typically determined during the system design phase as a trade-off between data collection efforts and accuracy, the former is determined by network deployment. Since our results in Sections IV-A showed a marked expansion in network deployment in 2023 compared to 2021, an increase in feature density can be expected as well. This is indeed the case for both Op_1^I and Op_2^I, as shown in Fig. 18 and Fig. 19, which present the number of unique features detected in each measurement point in a typical OD campaign in 2023 (right) vs. 2021 (left) for Op_1^I and Op_2^I, respectively (we do not report figures for Op_3^I and Op_4^I, but the following analyses

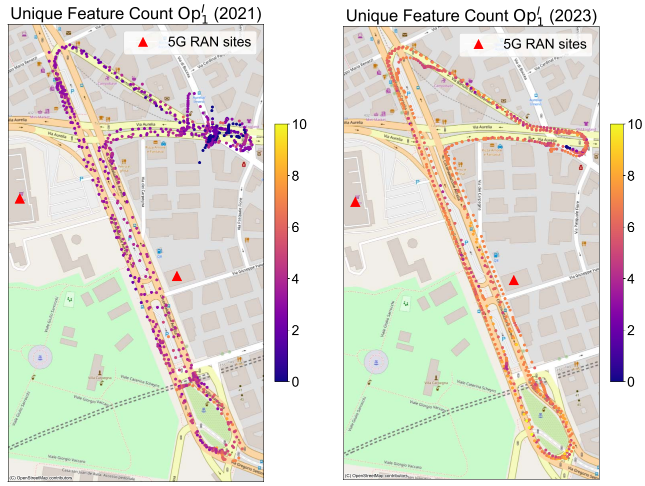


FIGURE 18. Number of unique features detected in each measurement point for a typical measurement campaign for Op_1^I in the OD location, in 2021 (left) vs. 2023 (right). The number of features is determined by all the detected combinations of PCI, SSB Index and NRAFCN for the MNC identifying Op_1^I.

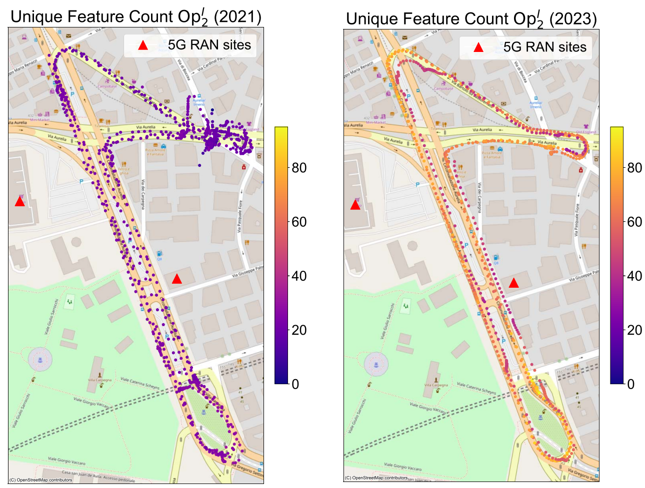


FIGURE 19. Number of unique features detected in each measurement point for a typical measurement campaign for Op_2^I in the OD location, in 2021 (left) vs. 2023 (right). The number of features is determined by all the detected combinations of PCI, SSB Index and NRAFCN for the MNC identifying Op_2^I.

also hold for these two MNOs).[2] The results show that the number of unique features has increased for both MNOs in 2023, as a result of their richer 5G deployments. The increase is larger for Op_2^I, thanks to the simultaneous increase in the number of PCIs and SSB carrier frequencies f_c; in the case of Op_1^I, a single f_c was detected in this location, consistent with the very limited use of multiple f_c observed in 2023 for this MNO (see Sections IV-A), leading to a moderate feature increase only due to a higher number of PCIs.

Fig. 20 (top) further highlights the variation in the number of features by comparing the distribution of features for the four Italian MNOs with a 5G RAN deployed in 2021 vs. 2023 in the common OD location. Fig. 20 (top) supports the

[2]Note that the two figures use different ranges for the number of unique features, in order to ensure good visibility for both MNOs.

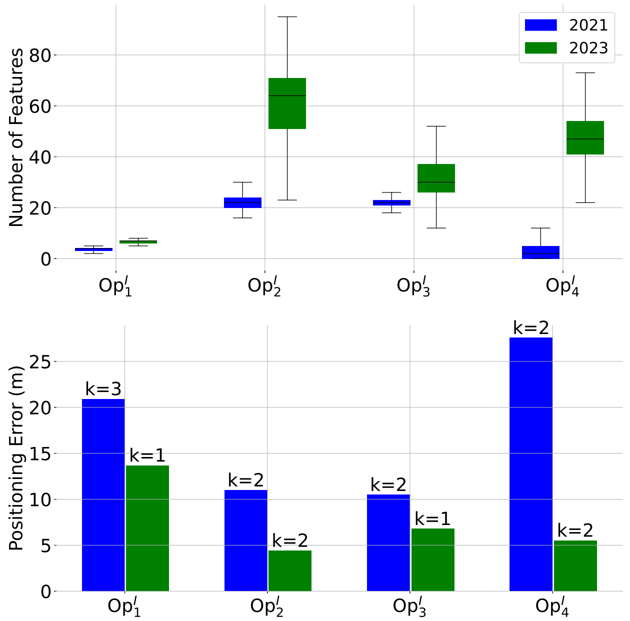


FIGURE 20. Distribution of features detected in the OD location (top) and Minimum Average Positioning Error and corresponding k value (bottom) for Italian MNOs in 2021 vs. 2023.

spatial-temporal analysis presented in Fig. 18 and Fig. 19 for Op_1^I and Op_2^I: 2023 data are characterized by an increase in the median value and variability of features for all MNOs compared to 2021; the increase is larger for the three MNOs that were detected on multiple f_c (vs. the single f_c observed in 2021) than for Op_1^I, still using a single f_c in the OD location under test.

The impact of the increased feature density on positioning accuracy is presented in Fig. 20 (bottom), showing the Minimum Average Positioning Error (and the corresponding k value) for the four MNOs for 2021 vs. 2023; results were obtained by averaging over 1000 runs, where in each run data were randomly divided in 800 RPs vs. 392 TPs. Results show that the increased feature density leads to better positioning accuracy for all MNOs, with a reduction of the Minimum Average Positioning Error between 20% and 80%. Fig. 20 highlights an inverse correlation between the feature density and the positioning error for both 2021 and 2023. Op_2^I and Op_4^I, in particular, benefited from the increased feature density, reaching a positioning error of about 5 meters; for Op_4^I, however, this was mostly due to a poor deployment in 2021, as shown by the low median number of features.

The next analysis is focused on Op_2^I in order to address both the impact of the introduction of multiple frequencies in the deployment, and the role of spatial density. Fig. 21 presents the Minimum Average Positioning Error as a function of the number of RPs (while keeping the number of TPs set at 392) in three cases: 2021, 2023 using all detected frequencies, and 2023 using only the data collected in the same single f_c that was detected in 2021 (out of the 6 f_c used by this MNO in 2023). Two observations can be drawn from the results: first, both richer PCI deployment and

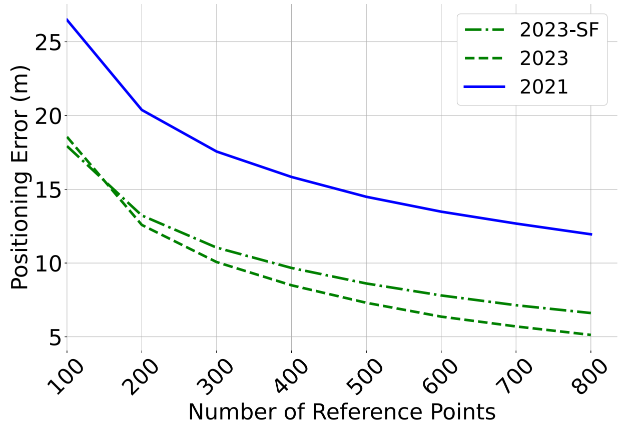


FIGURE 21. Minimum Average Positioning Error as a function of the number of RPs for Op_2^I in 2021, in 2023 using the single f_c detected in 2021 (2023-SF), and in 2023 using all detected f_c (2023).

multiple carrier frequencies contribute to the improvement in positioning accuracy; second, the same accuracy can be achieved by different combinations of feature density and spatial density, opening the way to different trade-offs and to the introduction of strategies to adapt spatial density (and thus data collection efforts) to network deployment characteristics, given a target positioning accuracy.

VIII. CONCLUSION

In this paper, we provide an in-depth empirical investigation of 5G NSA networks from deployment, performance, and service evolution perspectives. By leveraging our multi-year, cross-country, and multi-MNO dataset, open-sourced with the paper, we analyze several aspects. First, we observe a clear decrease of DL throughput performance and real-time gaming QoS/QoE over time, although MNOs are actively densifying their 5G networks in both physical and frequency domains. We identify worse radio coverage and 5G connectivity issues as root causes, at least partially caused by more aggressive policies in the use of 5G RAN by MNOs. We also observe gains in the UL throughput of one MNO, which now uses 4G-5G DC more efficiently. We then expand our analysis of interactive services and perform, similar to the real-time gaming case, a systematic assessment of AR/VR services. We show that maintaining stable 5G connectivity is crucial for these services, with data-intensive AR/VR applications posing increasing challenges resulting in poor QoE, ultimately suggesting the necessity for alternative architectural solutions (e.g., based on edge deployments). We also analyze international roaming and reveal clear detrimental effects on the QoE of interactive services, as a result of higher latency due to home-routed roaming and 4G fallback. Additionally, we provide a dedicated analysis on HO and connectivity management, which highlights how the configurations of such functions may significantly contribute to user-experienced 5G connectivity and corresponding performance. Finally, we analyze 5G-based positioning, a key enabler for several over-the-top applications, and show that a fingerprinting-based approach is able to provide quite accurate position estimates, with errors as low as 5 meters, benefiting from the dense and multi-frequency 5G network deployments observed in 2023. Our analyses show key insights on current 5G systems, thus paving the way towards better understanding and new studies aiming to enhance mobile systems beyond 5G.

REFERENCES

[1] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang, and N. Li, "5G deployment: Standalone vs. non-standalone from the operator perspective," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 83–89, Nov. 2020.

[2] (Statista, Hamburg, Germany). *5G Core Deployments Worldwide from 2018 to 2023, by Type*. 2022, Accessed: Jul. 2024. [Online]. Available: https://www.statista.com/statistics/1330511/5g-core-deployments-worldwideby-type/

[3] (GSMA, London, U.K.). *Second Wave of 5G: 30 Countries to Launch Services in 2023*. 2023, Accessed: Jul. 2024. [Online]. Available: https://www.gsma.com/newsroom/press-release/second-wave-of-5g-30-countries-to-launch-services-in-2023/

[4] K. Kousias et al., "A large-scale dataset of 4G, NB-IoT, and 5G non-standalone network measurements," *IEEE Commun. Mag.*, vol. 62, no. 5, pp. 44–49, May 2024, doi: 10.1109/MCOM.011.2200707.

[5] F. Mogyorósi et al., "Positioning in 5G and 6G networks—A survey," *Sensors*, vol. 22, no. 13, p. 4757, 2022.

[6] A. Narayanan et al., "A first look at commercial 5G performance on smartphones," in *Proc. Web Conf.*, 2020, pp. 894–905.

[7] A. Narayanan et al., "A variegated look at 5G in the wild: Performance, power, and QoE implications," in *Proc. ACM SIGCOMM Conf.*, 2021, pp. 610–625.

[8] D. Xu et al., "Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption," in *Proc. ACM SIGCOMM Conf. Spec. Interest Group Data Commun. Appl., Technol., Archit., Protocols Comput. Commun.*, 2020, pp. 479–494.

[9] K. Kousias et al., "Coverage and performance analysis of 5G non-standalone deployments," in *Proc. 16th ACM MobiCom Workshop Wireless Netw. Testbeds, Exp. Eval. Charact.*, 2022, pp. 61–68, doi: 10.1145/3556564.3558233.

[10] K. Kousias et al., "Implications of handover events in commercial 5G non-Standalone deployments in Rome," in *Proc. ACM SIGCOMM Workshop 5G Beyond Netw. Meas., Model., Use Cases*, 2022, pp. 22–27, doi: 10.1145/3538394.3546041.

[11] K. Kousias et al., "Empirical performance analysis and ML-based modeling of 5G non-standalone networks," *Comput. Netw.*, vol. 241, Mar. 2024, Art. no. 110207, doi: 10.1016/j.comnet.2024.110207.

[12] G. Caso et al., "An initial look into the performance evolution of 5G non-standalone networks," in *Proc. 7th Netw. Traffic Meas. Anal. Conf. (TMA)*, 2023, pp. 1–4, doi: 10.23919/TMA58422.2023.10199039.

[13] U. Ali et al., "Large-scale dataset for the analysis of outdoor-to-indoor propagation for 5G mid-band operational networks," *Data*, vol. 7, no. 3, p. 34, 2022, doi: 10.3390/data7030034.

[14] U. Ali et al., "Data-driven analysis of outdoor-to-indoor propagation for 5G mid-band operational networks," *Future Internet*, vol. 14, no. 8, p. 239, 2022, doi: 10.3390/fi14080239.

[15] W. Ye, X. Hu, T. Liu, R. Sun, Y. Li, and Z.-L. Zhang, "5GNN: Extrapolating 5G measurements through GNNs," in *Proc. 1st Int. Workshop Graph Neural Netw.*, 2022, pp. 36–41.

[16] R. A. K. Fezeu et al., "An in-depth measurement analysis of 5G mmwave PHY latency and its impact on end-to-end delay," in *Proc. Int. Conf. Passive Act. Netw. Meas.*, 2023, pp. 284–312.

[17] A. Narayanan et al., "Lumos5G: Mapping and predicting commercial mmWave 5G throughput," in *Proc. ACM Internet Meas. Conf.*, 2020, pp. 176–193.

[18] A. Hassan et al., "Vivisecting mobility management in 5G cellular networks," in *Proc. ACM SIGCOMM Conf.*, 2022, pp. 86–100.

[19] Z. Zhang, Y. Liu, Q. Li, Z. Liu, C. Peng, and S. Lu, "Dependent misconfigurations in 5G/4.5G radio resource control," *Proc. ACM Netw.*, vol. 1, pp. 1–20, Jul. 2023.

[20] C. Fiandrino, D. J. Martínez-Villanueva, and J. Widmer, "A study on 5G performance and fast conditional handover for public transit systems," *Comput. Commun.*, vol. 209, pp. 499–512, Sep. 2023.

[21] E. Ramadan, A. Narayanan, U. K. Dayalan, R. A. K. Fezeu, F. Qian, and Z.-L. Zhang, "Case for 5G-aware video streaming applications," in *Proc. ACM SIGCOMM Workshop 5G Meas., Model., Use Cases (5G-MeMU)*, 2021, pp. 27–34.

[22] A. A. El-Saleh, M. A. Al Jahdhami, A. Alhammadi, Z. A. Shamsan, I. Shayea, and W. H. Hassan, "Measurements and analyses of 4G/5G mobile broadband networks: An overview and a case study," *Wireless Commun. Mobile Comput.*, vol. 2023, no. 1, 2023, Art. no. 6205689. [Online]. Available: https://dl.acm.org/doi/abs/10.1155/2023/6205689

[23] M. I. Rochman et al., "A comprehensive analysis of the coverage and performance of 4G and 5G deployments," *Comput. Netw.*, vol. 237, Dec. 2023, Art. no. 110060.

[24] W. Ye, J. Carpenter, Z. Zhang, R. A. K. Fezeu, F. Qian, and Z.-L. Zhang, "A closer look at stand-alone 5G deployments from the UE perspective," in *Proc. IEEE Int. Mediterr. Conf. Commun. Netw. (MeditCom)*, 2023, pp. 86–91.

[25] Y. Liu and C. Peng, "A close look at 5G in the wild: Unrealized potentials and implications," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2023, pp. 1–10.

[26] M. Ghoshal et al., "Performance of cellular networks on the wheels," in *Proc. ACM Internet Meas. Conf.*, 2023, pp. 678–695.

[27] P. Parastar, A. Lutu, Ö. Alay, G. Caso, and D. Perino, "Spotlight on 5G: Performance, device evolution and challenges from a mobile operator perspective," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2023, pp. 1–10.

[28] X. Yuan et al., "Understanding 5G performance for real-world services: A content provider's perspective," in *Proc. ACM SIGCOMM Conf.*, 2022, pp. 101–113.

[29] *QualiPoc Android—The Premium Handheld Troubleshooter*, Rohde Schwarz, Munich, Germany, 2023, Accessed: Jul. 2024. [Online]. Available: www.rohde-schwarz.com/us/products/test-and-measurement/network-data-collection/qualipoc-android_63493-55430.html

[30] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. Cambridge, MA, USA: Academic, 2020.

[31] "Speedtest." Ookla, 2006, 2023, Accessed: Jul. 2024. [Online]. Available: https://www.speedtest.net

[32] "Speech and multimedia transmission quality (STQ); best practices for robust network QoS benchmark testing and scoring; Version 1.1.1," ETSI, Sophia Antipolis, France, Rep. TR 103 559, 2019.

[33] "Speech and multimedia transmission quality (STQ); QoS parameter measurements based on fixed data transfer times; Version 1.2.1," ETSI, Sophia Antipolis, France, Rep. TR 102 678, 2011.

[34] "Speech and multimedia transmission quality (STQ); QoS parameters and test scenarios for assessing network capabilities in 5G performance measurements; Version 1.1.1," ETSI, Sophia Antipolis, France, Rep. TR 103 702, 2020.

[35] *Latency Measurement and Interactivity Scoring Under Real Application Data Traffic Patterns*, ITU-Rec.-T G.1051, Int. Telecommun. Union, Geneva, Switzerland, 2023.

[36] (Rohde Schwarz, Munich, Germany). *Interactivity Test: QoS/QoE Measurements in 5G (Part 1)*. 2020, Accessed: Jul. 2024. [Online]. Available: https://www.rohde-schwarz.com/us/solutions/test-and-measurement/mobile-network-testing/stories-insights/article-interactivity-test-qos-qoe-measurements-in-5g-part-1-_253244.html

[37] "A two-way active measurement protocol (TWAMP)," Internet Eng. Task Force, RFC 5357, 2008, Accessed: Jul. 2024. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc5357

[38] *5G; System Architecture for the 5G System, Release 15*, 3GPP Standard TS 23.501, 2018.

[39] "Study on communication for automation in vertical domains (CAV)," 3GPP, Sophia Antipolis, France, Rep. TR 22.804, Oct. 2017.

[40] A. M. Mandalari et al., "Measuring roaming in Europe: Infrastructure and implications on users' QoE," *IEEE Trans. Mobile Comput.*, vol. 21, no. 10, pp. 3687–3699, Oct. 2022.

[41] *5G; NR; Radio Resource Control (RRC); Protocol Specification, Release 15, Version 15.15.0*, ETSI Standard TS 138 331, 2021.

[42] G. Caso, L. D. Nardis, F. Lemic, V. Handziski, A. Wolisz, and M.-G. D. Benedetto, "ViFi: Virtual fingerprinting WiFi-based indoor positioning via multi-wall multi-floor propagation model," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1478–1491, Jun. 2020, doi: 10.1109/TMC.2019.2908865.

GIUSEPPE CASO (Member, IEEE) received the B.Sc. degree in information and communication engineering from the Federico II University of Naples, Naples, Italy, and the M.Sc. and Ph.D. degrees in communication engineering from the Sapienza University of Rome, Rome, Italy. He is a Senior Lecturer with the Department of Mathematics and Computer Science, Karlstad University. He was previously with Ericsson Research, Kista, Sweden; Simula Metropolitan, Oslo, Norway; and the Sapienza University of Rome. He held various visiting positions with the Leibniz University of Hannover, Hannover, Germany; King's College London, London, U.K.; and the Technical University of Berlin, Berlin, Germany. His research interests include wireless communications, cognitive radio networks, mobile and distributed intelligent systems, IoT technologies, and location-based services.

MOHAMMAD RAJIULLAH received the B.Sc. degree in computer science from the Islamic University of Technology, Bangladesh, the M.Sc. degree in global information and telecommunication studies from Waseda University, Japan, and the Ph.D. degree in computer science from Karlstad University, Sweden, where he is a Senior Lecturer. His research interests are in the areas of low latency networking, Web performance, mobile networks, and IoT.

KONSTANTINOS KOUSIAS received the B.Sc. and M.Sc. degrees from the Department of Electrical and Computer Engineering, University of Thessaly, Greece, and the Ph.D. degree from the Faculty of Mathematics and Natural Sciences, University of Oslo, where he is a Postdoctoral Researcher with the Department of Informatics. His research focuses on the empirical modeling and evaluation of mobile and IoT systems using data analytics and AI.

USMAN ALI (Student Member, IEEE) received the B.Sc. degree from the Federal Urdu University of Arts, Sciences and Technology, Islamabad, Pakistan, in 2015, the M.Sc. degree from COMSATS University, Islamabad, Pakistan, in 2018, and the Ph.D. degree in information and communications technology from the Sapienza University of Rome, Italy, in 2024. His current research interests include multiple antenna communications, limited feedback techniques, and rate-splitting multiple access.

NADIR BOUZAR is currently pursuing the Ph.D. degree with the Sapienza University of Rome. His research interests include the design and evaluation of user/device positioning schemes using UWB and cellular technologies.

LUCA DE NARDIS (Member, IEEE) received the Ph.D. degree from the Sapienza University of Rome, Italy, in 2005, where he is currently an Associate Professor with the Department of Information Engineering, Electronics and Telecommunications. In 2007, he was a Postdoctoral Fellow with the University of California at Berkeley, Berkeley. He authored or coauthored over 120 international peer-reviewed publications. His research interests include cognitive communications, medium access control, routing protocols, and wireless positioning systems.

ANNA BRUNSTROM (Member, IEEE) received the B.Sc. degree in computer science and mathematics from Pepperdine University, Malibu, CA, USA, in 1991, and the M.Sc. and Ph.D. degrees in computer science from the College of William and Mary, Williamsburg, VA, USA, in 1993 and 1996, respectively. She is a Full Professor and a Research Manager with the Distributed Intelligent Systems and Communications Research Group, Karlstad University and a part-time Distinguished Researcher with the Institute of Software Engineering and Technologies, University of Malaga, Spain. She has authored/coauthored over 200 international peer-reviewed publications. She has served as a Principal Investigator and a Coordinator with Karlstad University in several national and international projects. Her research interests include transport protocol design, techniques for low-latency Internet communication, multi-path communication, and performance evaluation of mobile systems, including 5G/6G.

OZGU ALAY (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and electronic engineering from Middle East Technical University, Turkey, and the Ph.D. degree in electrical and computer engineering from the Tandon School of Engineering, New York University. She is a Full Professor with the University of Oslo and Adjunct Professor with Karlstad University. She is the author of more than 70 peer-reviewed IEEE and ACM publications. She actively serves on technical boards of major conferences and journals, and has participated as principal investigator in numerous EU-funded research projects in the area of mobile cellular systems. Her research interests include mobile networks, low-latency networking, multipath protocols, and multimedia transmission over wireless networks.

MARCO NERI received the B.Sc. and M.Sc. degrees in electrical engineering from the Sapienza University of Rome, Rome, Italy, in 2015 and 2017, respectively, and the master's degree in business administration from the LUISS Business School, Rome, in 2023. In 2017, he joined Rohde & Schwarz, where he is currently a Staff Application Engineer for the Mobile Network Testing unit. His research interests include measurement-based and data-driven analyses of 5G and cellular IoT systems.

MARIA-GABRIELLA DI BENEDETTO (Fellow, IEEE) received the Ph.D. degree from the Sapienza University of Rome, Italy, in 1987. In 1991, she joined the Faculty of Engineering, Sapienza University of Rome, where she is currently a Full Professor of Telecommunications. She held various visiting positions with the Massachusetts Institute of Technology; the University of California at Berkeley, Berkeley; and the University of Paris-Saclay. She is currently a Research Affiliate with the Research Laboratory of Electronics, MIT. She is a Fellow of the Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA, USA. Her research interests include wireless communication systems, impulse radio communications, and speech. In 1994, she received the Mac Kay Professorship Award from the University of California at Berkeley.